THE UNIVERSITY OF ALBERTA

UNBIASED ESTIMATION, SIMPLICITY AND OPTIMALITY IN CERTAIN BINOMIAL SAMPLING PLANS

by

IHOR ZINOVIE CHORNEYKO

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

EDMONTON, ALBERTA

JULY, 1965

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Unbiased Estimation, Simplicity and Optimality in Certain Binomial Sampling Plans" submitted by Ihor Zinovie Chorneyko in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

ACKNOWLEDGEMENT

The author wishes to acknowledge the guidance of Dr. T.V. Narayana without whose kind efforts and gentle persuasion, this thesis would not have been possible.

ABSTRACT

The purpose of this thesis is to investigate two classes of problems with reference to binomial sampling plans. Firstly, we make a detailed examination of simple binomial sampling plans with two parameters from the point of view of estimation theory, studying, in particular, the completeness of such sampling plans. The second problem with which we are concerned is the role of simplicity in the optimal truncated sequential plans described by Blackwell and Girshick. In this connection, we study the relationship of binomial sampling plans as described by Girshick, Mosteller and Savage and the "optimal plans" described by Blackwell and Girshick.

Chapter I presents a resume of the relevant work on binomial sampling plans. We reformulate several of the results and present a different approach and some alternative definitions to bring out more clearly the fundamental problems of unbiased sequential estimation for two parameter binomial populations. We illustrate these results by proving the completeness of two families of two parameter binomial distributions and list explicitly the polynomials estimable unbiasedly in these plans. The optimal sampling plans of Blackwell and Girshick are interpreted in terms of the more intuitive characterization of binomial sampling plans first given by Girshick, Mosteller and Savage as a preliminary to our discussion in Chapter III, which enables us to decide when an optimal plan is simple.

In Chapter II, we prove the completeness of another family of two parameter binomial distributions. Using the concepts and alternative

definitions given in Chapter I, we are able to give the number of polynomials in the basis of the linear space of estimable polynomials for large classes of simple, sampling plans and an upper bound for this number for any simple, sampling plan. Certain topics and theorems which have been touched upon but not fully discussed in the literature are brought out as a natural consequence of our discussion.

The relationship between simplicity and the risk functions of optimal sampling plans is discussed in Chapter III. By means of an algorithm, we give a method for discerning readily whether the sampling plans described by the cylinder sets of Blackwell and Girshick are simple or not. Sufficient conditions for certain optimal sampling plans to be simple are given and illustrated by several examples.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

SIMPLE SAMPLING PLANS AND SOME OF THEIR PROPERTIES

1.1 Introduction

Binomial sampling plans and problems of estimation and completeness for these plans were first studied by Girshick, Mosteller and Savage [9] in 1946. They described these plans in terms of boundary, continuation and inaccessible points and introduced the concept of simplicity. Further work on simple sampling plans was carried out by De Groot [6], who studied unbiased sequential estimation in these plans and gave an elegant characterization of simple sampling plans of size n. Other characterizations of such plans in terms of vectors with non-negative integral components are available [(5), (12)] and are due to Narayana, Brainerd and Mohanty. Simple sampling plans with two parameters, which can be considered as an extension of the basic model in [9], have occasionally been considered in the literature e.g. Gabriel [8] and Blackwell and Girshick [4], (cf. p. 222). We introduce and study in this chapter the simplest sampling plans of size n with 2 parameters in order to indicate the type of estimation problems considered in this thesis.

From a very different point of view, Blackwell and Girshick have delineated "optimal sequential sampling plans" in their discussion of Bayes procedures for sequential games ([4], particularly Chapters 9 and 10). While it is easily seen that their very general definition of sequential sampling plans in terms of cylinder sets includes all binomial sampling plans of size n, we propose to investigate optimal sequential sampling plans in certain special cases where the definitions in terms of cylinder

sets and in terms of boundary, continuation and inaccessible points are both available. With this end in view, we shall present a brief though somewhat incomplete review of the results of Blackwell and Girshick. It is not the purpose of this chapter to provide a detailed resumé of all the diverse results known concerning sampling plans, since excellent accounts are available in the references quoted above, ([4], [5], [6], [12]). We shall however, present a review of certain results and theorems which are directly relevant to our work, and especially indicate various modifications of notations and results which will prove useful in later chapters.

Section 1.2 describes the simplest sampling plan of size n with 2 parameters, namely the "fixed" sampling plan of size n. The next two sections discuss various estimation and completeness properties of this plan and motivate our study of simple sampling plans with 2 parameters in Chapter 2 and the latter part of this thesis.

Sections 1.5 and 1.6 give a brief resumé of known results concerning characterizations of simple sampling plans and introduces the concept of "canonical sequences of deformations", following the development of Brainerd and Narayana [5]. The last section summarizes the procedure described by Blackwell and Girshick [4] for obtaining optimal sequential sampling plans and interprets this procedure in terms of boundary, continuation and inaccessible points in the many cases where this is possible.

1.2 The "Fixed" Sampling Plan of Size n.

In order to motivate our study of two parameter binomial sampling plans, we shall introduce in this section a particularly simple plan of

this type, namely the "fixed" sampling plan of size n. This sampling plan generates a distribution which we shall call A_n. For the sake of simplicity, we shall use the symbol A_n to denote both the sampling plan and the distribution generated by this sampling plan. This distribution has already been studied by Gabriel [8].

All the plans considered in this thesis can be derived from the following model. Successive observations are made on chance variables $X_1, X_2, X_3, \ldots$ such that

$$P(X_1 = 1) = p_1, \quad P(X_1 = 0) = 1 - p_1 = q_1 \quad \text{and}$$

1.2.1 $$P(X_n = 1|X_{n-1} = 1) = p_2, \quad P(X_n = 0|X_{n-1} = 1) = 1 - p_2 = q_2$$

$$P(X_n = 1|X_{n-1} = 0) = p_1, \quad P(X_n = 0|X_{n-1} = 0) = 1 - p_1 = q_1,$$

where $0 < p_1 < 1$, $0 < p_2 < 1$ and $p_1 \neq p_2$. If $p_1 = p_2$, we have the usual one parameter binomial sampling plan. We can also consider this model as the following coin tossing problem. Two coins, 1 and 2, are tossed with probabilities p_1 and p_2, $0 < p_1 < 1$, $0 < p_2 < 1$, $p_1 \neq p_2$, of falling heads. On the first toss, coin 1 is used. On the j^{th} toss, coin 1 is used if the j-1st toss was a tail, and coin 2 is used if the j-1st toss was a head. By specifying various stopping rules, we can obtain different distributions (sampling plans). We shall consider several such stopping rules in the next sections and in Chapter II.

<u>Definition 1.2.1</u>. The distribution obtained by stopping as soon as n tosses are made is called A_n.

By using coin 2 at the first toss and then proceeding as above, we obtain the distribution which we call A'_n. It is easily seen that the

distributions A_n and A'_n are generated by the "fixed" sampling plan of size n under the model 1.2.1. In De Groot's notation [6], the boundary, B, for A_n and A'_n consists of all those points $\gamma = (x,y)$ where x and y are non-negative integers such that $x + y = n$, $n > 0$.

In terms of the model 1.2.1, we can determine the probability of reaching any point $\gamma = (x,y)$. In the usual way, we can associate the sequence of 1's and 0's with a lattice path in the plane, where each 1 is represented by a unit horizontal step and each 0 by a unit vertical step. Since we shall be dealing with such lattice paths only, we shall simply call them paths. Let k be the number of right angle turns of the type $\lrcorner$ (k is the number of (1,0) joins) in the path to the point $\gamma = (x,y)$. The probability associated with this path is easily found to be

$$p_1^{k+1} \; q_1^{y-k} \; p_2^{x-k-1} \; q_2^{k} \quad \text{if the last step is horizontal (1),}$$

and

$$p_1^{k} \; q_1^{y-k} \; p_2^{x-k} \; q_2^{k} \quad \text{if the last step is vertical (0).}$$

The probability $p(\gamma)$, of reaching the point $\gamma = (x,y)$ is

$$1.2.2 \qquad p(\gamma) = \sum_{k \in K} N(\gamma,k)\, p_1^{k+1} q_1^{y-k} p_2^{x-k-1} q_2^{k} + \sum_{k \in K'} N'(\gamma,k)\, p_1^{k} q_1^{y-k} p_2^{x-k} q_2^{k},$$

where $N(\gamma,k)$ and $N'(\gamma,k)$ represent the number of paths to the point $\gamma = (x,y)$ with $k(1,0)$ joins and ending in a 1 or a 0 respectively. The index set K is a subset of the set of integers $0,1,2,\ldots \min(x-1,y)$ and K' is a subset of the integers $0,1,2,\ldots,\min(x,y)$. For any stopping rule

(sampling plan), we can completely determine the distribution by determining K, K', $N(\gamma,k)$ and $N'(\gamma,k)$ for each γ in the boundary. Notice that equation 1.2.2 is the obvious two parameter analogue of the one parameter case (c.f. De Groot [6], p. 82).

1.3 Completeness and Estimation in the Distributions A_n, A'_n.

In this section, we shall use the method of generating functions to prove the completeness of the distributions A_n and A'_n. This method has the advantage that it enables us to list the polynomials estimable unbiasedly. Then, by using the standard Rao-Blackwell procedure, we obtain the unique, unbiased, minimum variance estimator of p_i.

Let us suppose we toss the coins n times using coin 1 on the first toss (A_n). Let r be the number of heads obtained. Let $p_{n,r}$ be the probability of obtaining r heads in n tosses. Then, according to 1.2.2

1.3.1
$$\begin{aligned} p_{n,r} &= \sum_{k \in K} N(\gamma,k)\, p_1^{k+1} q_1^{n-r-k} p_2^{r-k-1} q_2^{k} + \\ &\quad \sum_{k \in K'} N'(\gamma,k)\, p_1^{k} q_1^{n-r-k} p_2^{r-k} q_2^{k} \\ &= \sum_{k=0}^{r-1} \binom{n-r}{k}\binom{r-1}{k} p_1^{k+1} q_1^{n-r-k} p_2^{r-k-1} q_2^{k} + \\ &\quad \sum_{k=1}^{r} \binom{n-r}{k}\binom{r-1}{k-1} p_1^{k} q_1^{n-r-k} p_2^{r-k} q_2^{k} . \end{aligned}$$

The coefficients $N(\gamma,k)$ are obtained as follows. Since there are $k(1,0)$ joins, this implies that among the $r - 1$ heads occurring before the last

head there will be k changes to tails. These changes may occur in $\binom{r-1}{k}$ ways. Also, these k changes occur before tails, which may be arranged in $\binom{n-r}{k}$ ways. The total number of ways in which $k(1,0)$ joins may appear is $N(\gamma,k) = \binom{r-1}{k}\binom{n-r}{k}$. A similar argument yields $N'(\gamma,k)$.

If we toss the coin n times using coin 2 at the first toss, we can obtain a similar expression, $p'_{n,r}$ for the probability of obtaining r heads in n tosses in the game A'_n, i.e.

1.3.2
$$p'_{n,r} = \sum_{k=1}^{r} \binom{n-r-1}{k-1}\binom{r}{k} q_1^{n-r-k} p_1^k q_2^k p_2^{r-k}$$
$$+ \sum_{k=0}^{r} \binom{n-r-1}{k}\binom{r}{k} q_1^{n-r-k-1} p_1^k q_2^{k+1} p_2^{r-k}$$

It is clear that by putting $p_1 = p_2$, we obtain the usual one parameter binomial distribution.

In what follows, we shall use the ideas and notation developed in Blackwell and Girshick [4] Chapter 3 and 8, to describe sample spaces, sufficient partitions and sufficient statistics. Let $\mathcal{Z} = (Z, \Omega, p_w)$, where Z consists of points z representing outcomes in A_n, $\Omega = \left\{(p_1,p_2)\{p_1 : 0 < p_1 < 1\} \times \{p_2 : 0 < p_2 < 1\}, p_1 \neq p_2\right\}$ and $p_w(z)$ is the probability of $z \in Z$ given $w \in \Omega$.

Let us consider the partition S of Z where $S = (s^1_{10} \ldots s^1_{rk} \ldots s^1_{no}, s^2_{oo} \ldots s^2_{mk} \ldots s^2_{n-1,1})$. The superscripts 1 and 2 indicate whether the last step in the sequence z was a head or tail and the subscripts indicate the number of heads and the number of right angle turns respectively. It is easily seen that S is a sufficient

partition. Let $Y = (R,k)$, where $R = \pm r$ according as the sequence ends with a head or a tail. The number of successes is $r = |R|$. Thus, Y is seen to be a sufficient statistic.

Lemma 1.3.1 Let $A_n(s)$, $A'_n(s)$ represent the generating functions for the probability distributions A_n and A'_n. Then

$$A_n(s) = q_1 \; A_{n-1}(s) + p_1 s \; A'_{n-1}(s).$$

$$A'_n(s) = q_2 \; A_{n-1}(s) + p_2 s \; A'_{n-1}(s).$$

Proof: We shall give the proof for $A_n(s)$ only, as the proof for $A'_n(s)$ is similar. The proof follows from the fact that after the first toss the resulting game is A_{n-1} or A'_{n-1} depending on whether the first toss resulted in a tail or a head.

Lemma 1.3.2 The number of distinct values taken on by the sufficient statistic Y (with positive probability) for A_n is $\frac{n(n+1)}{2} + 1$.

Proof: Suppose n is odd. We count the values of Y (taken on with positive probability) as follows:

1. Let $R > 0$. Then each sequence ends with a head and for a given R the possible values of k range from 0 to $\min(R - 1, n - R)$.

2. $R = 0$. The only value of k is 0.

3. $R < 0$. Each sequence ends with a tail and k cannot be zero. For a given R, k ranges from 1 to $\min(|R|, n - |R|)$. Thus the number of values taken on by Y is

$$(1 + 2 + \dots + \frac{n+1}{2} + \frac{n+1}{2} - 1 + \dots + 2 + 1) + (1) +$$

$$(1 + 2 + \dots + \frac{n-1}{2} + \frac{n-1}{2} + \frac{n-3}{2} + \dots + 1) = \frac{n(n+1)}{2} + 1;$$

where, for example, the first bracket indicates the distinct values of Y in case (1). A similar counting procedure yields the same result if n is even.

<u>Theorem 1.3.1</u> The distributions A_n, A'_n are complete.

<u>Proof</u>: Clearly the distributions A_1, A'_1 are complete. We shall follow the idea expressed in [6] p. 97 to show A_n is complete. We shall prove by induction that any maximal set of linearly independent polynomials in p_1 and p_2 which are estimable unbiasedly contain $\frac{n(n + 1)}{2} + 1$ elements, thus showing that A_n is complete. For A_4, such a set is

$$
\begin{array}{lllll}
1 & p_1 & p_1^2 & p_1^3 & p_1^4 \\
p_1p_2 & p_1p_2^2 & p_1p_2^3 & & \\
p_1^2p_2 & p_1^2p_2^2 & & & \\
p_1^3p_2. & & & &
\end{array}
$$

It is clear that such a set is not unique. There are many other sets of linearly independent polynomials spanning the space of estimable polynomials. Another such equivalent set for A_4 is

$$
\begin{array}{lllll}
1 & p_1 & p_1^2 & p_1^3 & p_1^4 \\
p_1q_2 & p_1q_2^2 & p_1q_2^3 & & \\
p_1^2q_2 & p_1^2q_2^2 & & & \\
p_1^3q_2. & & & &
\end{array}
$$

We shall now assume

1. that the distributions A_n, A'_n are complete for $n \leq k$.

2. that the basis for the linear space of estimable polynomials in A_k, A'_k can be taken to be

$$A_k$$

$$\text{1.3.3} \qquad \begin{array}{lllll}
1 & p_1 & p_1^2 & \cdots & p_1^k \\
p_1p_2 & p_1p_2^2 & \cdots & p_1p_2^{k-1} & \\
p_1^2p_2 & p_1^2p_2^2 & \cdots & p_1^2p_2^{k-2} & \\
\vdots & & & & \\
p_1^{k-2}p_2 & p_1^{k-2}p_2^2 & & & \\
p_1^{k-1}p_2. & & & &
\end{array}$$

$$A_k'$$

$$\text{1.3.4} \qquad \begin{array}{lllll}
1 & p_2 & p_2^2 & \cdots & p_2^k \\
p_1q_2 & \cdots & p_1^{k-1}q_2 & & \\
p_1q_2^2 & \cdots & p_1^{k-2}q_2^2 & & \\
\vdots & & & & \\
p_1q_2^{k-2} & p_1^2q_2^{k-2} & & & \\
p_1q_2^{k-1} & & & &
\end{array}$$

It is easily seen that the polynomials in these sets are linearly independent. Now apply the generating function relationship of Lemma 1.3.1, i.e. multiply 1.3.3 by q_1 and 1.3.4 by p_1. The resulting set contains all the polynomials estimable unbiasedly in A_{k+1}. From this set we wish to select the maximal linearly independent set of polynomials. We can obtain this maximal set as follows. q_1 appears when we multiply 1.3.3 by q_1 and p_1 appears when we multiply 1.3.4 by p_1. Since $q_1 + p_1 = 1$, we can replace q_1 by 1. p_1p_2, $p_1p_2^2 \ldots p_1p_2^k$ are obtained from 1.3.4 on multiplication by p_1 and remain unchanged. Since $p_1^{j-1} - q_1p_1^{j-1} = p_1^j$, we replace $q_1p_1^{j-1}$, $j = 2,3,\ldots,k+1$, by p_1^j. Similarly, we replace

$q_1 p_1^{j-1} p_2^{i}$ by $p_1^{j} p_2^{i}$, $j = 3, 4, \ldots, k$, $i = k - 2, k - 3, \ldots, 1$. Since $p_1^{j} q_2^{i} = p_1^{j} \left(1 - \binom{i}{1} p_2 \ldots (-1)^{i} \binom{i}{i} p_2^{i}\right)$, we can remove all the remaining polynomials to obtain the maximal linearly independent set 1.3.3 with $n = k + 1$. Since each set contains $\frac{(k+1)(k+2)}{2} + 1$ elements, the proof is complete.

As one would intuitively expect, p_2 is not estimable in A_n, since it is conceivable that all observations result in a tail and coin 2 is not used at all.

We can obtain the unique, minimum variance unbiased estimator $\hat{p}_1$ of p_1 as follows.

Let $X_1 = 1$ if the first toss is a head.

$= 0$ otherwise.

By the Rao-Blackwell theorem,

$$\begin{aligned}
\hat{p}_1 &= E(X_1 \mid R, k) \\
&= \frac{N^*(\gamma,k)}{N(\gamma,k)}, \quad k \in K \\
&= \frac{N^{*\prime}(\gamma,k)}{N'(\gamma,k)}, \quad k \in K',
\end{aligned}$$

where $N^*(\gamma,k)$ is the number of paths in each of the two cases beginning with a head. For $k \in K$, $N^*(\gamma,k) = \binom{n-r}{k}\binom{r-1}{k} - \binom{n-r-1}{k}\binom{r-1}{k}$. For $k \in K'$, $N^{*\prime}(\gamma,k) = \binom{n-r}{k}\binom{r-1}{k-1} - \binom{n-r-1}{k}\binom{r-1}{k-1}$. Hence,

$$\begin{aligned}
\hat{p}_1 &= \frac{\binom{n-r}{k} - \binom{n-r-1}{k}}{\binom{n-r}{k}} = \frac{k}{n-r}, \quad r \neq n \\
&= 1 \qquad\qquad r = n, \text{ in both cases.}
\end{aligned}$$

1.4 Completeness and Estimation in A_{mn}, A'_{mn}.

In this section, we shall consider the distributions obtained by varying the stopping rule in the model described by 1.2.1.

Definition 1.4.1 Let us consider the model 1.2.1. If we stop whenever m ones (heads) or n zeroes (tails) appear, the resulting distribution is called A_{mn}. If initially $p(X_1 = 1) = p_2$, then the resulting distribution is called $A'_{m,n}$. Thus, the sampling plan which yields the distributions A_{mn}, A'_{mn} has boundary points $(m,0)$ $(m,1)$ ... $(m,n-1)$ $(0,n)(1,n)$ ... $(m-1,n)$. Let k be the number of $(1,0)$ joins. According to 1.2.2, the probability of reaching a boundary point (m,y) is given by

1.4.1 $$\sum_{k=0}^{\min(m-1,y)} \binom{y}{k}\binom{m-1}{k} p_1^{k+1} q_1^{y-k} p_2^{m-k-1} q_2^{k}$$

and that of reaching a boundary point (x,n) is given by

1.4.2 $$\sum_{k=1}^{\min(x,n)} \binom{n}{k}\binom{x-1}{k-1} p_1^{k} q_1^{n-k} p_2^{x-k} q_2^{k} .$$

If we let k be the number of $(0,1)$ joins in A'_{mn}, the probability of reaching a boundary point (m,y), $y \geq 1$, is

1.4.3 $$\sum_{k=1}^{\min(y,m)} \binom{y-1}{k-1}\binom{m}{k} p_1^{k} q_1^{y-k} p_2^{m-k} q_2^{k}$$

and the probability of reaching the point (x,n) is

1.4.4 $$\sum_{k=0}^{\min(x,n-1)} \binom{n-1}{k}\binom{x}{k} p_1^{k} q_1^{n-k-1} p_2^{x-k} q_2^{k+1} .$$

We shall use the method of generating functions developed in 1.3 to prove that the distributions A_{mn}, A'_{mn} are complete. Although this method cannot readily be extended to distributions determined by more general stopping rules, it does provide some insight into the nature of these distributions. As in 1.3, we can define a sufficient partition and a sufficient statistic $T = (N,k)$, where $N = x$ or $N = -y$ depending on whether the path is to the boundary point (x,n) or (m,y). The following lemmas determine the number of values T takes on with positive probability and the generating function relationship between the distributions.

Lemma 1.4.1 The sufficient statistic T takes on $mn + 1$ values with positive probability.

Proof: Let $m \geq n$. Since k ranges from 0 to $\min(m-1,y)$ and from 1 to $\min(x-1,n)$ the total number of values is $1 + 1 + 2 + 3 + \ldots n + (m - n - 1)n + 1 + 2 \ldots + n = mn + 1$.

If $m < n$, a similar counting procedure yields $1 + 1 + 2 + \ldots + m-1 + 1 + 2 + \ldots + (m-1) + \ldots (n-m + 1)m = mn + 1$.

Lemma 1.4.2 Let $A_{mn}(s)$ and $A'_{mn}(s)$ represent the generating functions for the probability distributions of A_{mn} and A'_{mn}. Then

$$A_{mn}(s) = q_1 \; A_{m,n-1}(s) + p_1 s \; A'_{m-1,n}(s).$$

$$A'_{mn}(x) = q_2 \; A_{m,n-1}(s) + p_2 s \; A'_{m-1,n}(s).$$

Proof: The proof follows from the fact that after the first trial, the remaining trials form a distribution for $A'_{m-1,n}$ and $A_{m,n-1}$ respectively.

The proof of Theorem 1.4.1 parallels that of Theorem 1.3.1, just as the two lemmas proven above are analogues of lemmas 1.3.1 and 1.3.2.

Theorem 1.4.1. The polynomials estimable unbiasedly in $A_{m,n}$ and $A'_{m,n}$ are linear combinations of the following sets of polynomials.

A_{mn}.

1.4.5
$$\begin{array}{ccccc} 1 & p_1 & p_1^2 & \cdots & p_1^n \\ p_1p_2 & p_1^2p_2 & \cdots & & p_1^np_2 \\ p_1p_2^2 & p_1^2p_2^2 & \cdots & & p_1^np_2^2 \\ \vdots & & & & \vdots \\ p_1p_2^{m-1} & p_1^2p_2^{m-1} & & & p_1^np_2^{m-1} \end{array}$$

A'_{mn}

1.4.6
$$\begin{array}{ccccc} 1 & q_2 & q_2^2 & \cdots & q_2^n \\ q_2q_1 & q_2^2q_1 & \cdots & & q_2^nq_1 \\ \vdots & & & & \vdots \\ q_2q_1^{m-1} & q_2^2q_1^{m-1} & \cdots & & q_2^nq_1^{m-1} \end{array}$$

Proof: First we consider the polynomials for the games A_{1n}, A_{m1}, A'_{1n}, A'_{m1}. For these games, the polynomials are:

(a) $p_1 \quad q_1p_1 \quad q_1^2p_1 \quad \cdots \quad q_1^{n-1}p_1 \quad q_1^n$

(b) $q_1 \quad p_1q_2 \quad p_1p_2q_2 \quad \cdots \quad p_1p_2^{m-2}q_2 \quad p_1p_2^{m-1}$

(c) $p_2 \quad q_2p_1 \quad q_2q_1p_1 \quad \cdots \quad q_2q_1^{n-2}p_1 \quad q_2q_1^{n-1}$

(d) $q_2 \quad p_2q_2 \quad p_2^{\,2}q_2 \quad \cdots \quad p_2^{\,m-1}q_2 \quad p_2^{\,m}$

These polynomials can be replaced by the equivalent sets

(a') $1 \quad p_1 \quad p_1^{\,2} \quad \cdots \quad p_1^{\,n}$ or $1 \quad q_1 \quad q_1^{\,2} \quad \cdots \quad q_1^{\,n}$

(b') $1 \quad p_1 \quad p_1p_2 \quad \cdots \quad p_1p_2^{\,m-2} \quad p_1p_2^{\,m-1}$

(c') $1 \quad q_2 \quad q_2q_1 \quad \cdots \quad q_2q_1^{\,n-2} \quad q_2q_1^{\,n-1}$

(d') $1 \quad p_2 \quad p_2^{\,2} \quad \cdots \quad p_2^{\,m}$ or $1 \quad q_2 \quad q_2^{\,2} \quad \cdots \quad q_2^{\,m}$

Thus, the theorem is true for A_{m1}, A_{1n}, A'_{m1}, A'_{1n}.

Proceeding by induction and using the generating function relationship, we obtain the following sets of polynomials.

(a)
$$\begin{array}{lllll} q_1 & q_1p_1 & q_1p_1^{\,2} & \cdots & q_1p_1^{\,n-1} \\ q_1p_1p_2 & \cdots & q_1p_1^{\,n-1}p_2 & & \\ \vdots & & & & \\ q_1p_1p_2^{\,m-1} & q_1p_1^{\,2}p_2^{\,m-1} & \cdots & q_1p_1^{\,n-1}p_2^{\,m-1} & \end{array}$$

and

(b)
$$\begin{array}{lllll} p_1 & p_1q_2 & p_1q_2^{\,2} & \cdots & p_1q_2^{\,n} \\ p_1q_2q_1 & p_1q_2^{\,2}q_1 & \cdots & p_1q_2^{\,n}q_1 & \\ \vdots & & & & \\ p_1q_2q_1^{\,m-2} & p_1q_2^{\,2}q_1^{\,m-2} & \cdots & p_1q_2^{\,n}q_1^{\,m-2} & \end{array}$$

Combine p_1 and q_1 and replace q_1 by 1. $p_1^{\,j}$, $j = 2,3,\ldots,n$, is obtained by combining $p_1^{\,j-1}$ and $q_1p_1^{\,j-1}$ and replacing $q_1p_1^{\,j-1}$ by $p_1^{\,j}$. $p_1^{\,j}p_2^{\,i}$, $j = 1,2,\ldots,n$, $i = 1,2,\ldots,m-1$ is obtained from $p_1^{\,j-1}p_2^{\,i}$

and $q_1p_1^{j-1}p_2^{i}$ and replacing $q_1p_1^{j-1}p_2^{i}$ by $p_1^{j}p_2^{i}$. The remaining polynomials $p_1q_2^{j}q_1^{i}$ $j = 1,2,\ldots,n$, $i = 1,2,\ldots,m-1$ are easily seen to be linear combinations of the polynomials already obtained and can be omitted. Thus, the maximal set of linearly independent polynomials is given by 1.4.5. By multiplying by q_2 and p_2 and going through a similar procedure as above, we obtain the maximal set 1.4.6.

Corollary. The sufficient statistic Y is complete.

Proof: The polynomials 1.4.5 and 1.4.6 span the space of estimable polynomials and are linearly independent. The polynomials given by 1.4.1, 1.4.2 and 1.4.3, 1.4.4 also span the space of estimable polynomials. Since the number of these polynomials is by Lemma 1.4.1 $mn + 1$, they must be linearly independent.

We conclude this section by pointing out that the results of 1.3 and 1.4 are a generalization of the one parameter case. We have already seen that putting $p_1 = p_2$ in 1.3.1 and 1.3.2 yields the one parameter binomial distribution. Since $\sum_{k=0}^{\min(x,z)} \binom{x}{k}\binom{z}{k} = \binom{x+z}{x}$, a similar one parameter distribution is obtained by putting $p_1 = p_2$ in 1.4.1 and 1.4.2. Also, De Groot [6] has proved that in a sampling plan of size n, all polynomials of degree at most n are estimable unbiasedly. But it is easily seen that putting $p_1 = p_2$ in 1.3.3 and 1.3.4 gives us the estimable polynomials of degree at most n. Similarly, if we put $p_1 = p_2$ in 1.4.5 and 1.4.6, we obtain all the estimable polynomials of degree at most $m + n - 1$. Since the size of $A_n(A'_n)$ is n, and the size of $A_{mn}(A'_{mn})$ is $m + n - 1$, we have obtained a direct generalization of De Groot's result for these cases.

1.5 Simple Sampling Plans and Vectors with Non-Negative Integral Components.

Simple sampling plans were first studied by Girshick, Mosteller and Savage [9] in connection with the estimation of the parameter in the binomial distribution. They defined a sampling plan to be simple if no two continuation points on the line $x + y = j$, $j = 0,1,2,\ldots$ are separated by boundary or inaccessible points. Formally, we state this definition as follows.

Definition 1.5.1 Let $m_j = \sum_{i=1}^{j} X_i$. A sampling plan is said to be simple, if and only if, for any two continuation points $(m_j, j-m_j)$ and $(m_j', j-m_j')$, the points $(m_j + 1, j - m_j - 1)$, $(m_j + 2, j - m_j - 2) \ldots$ $(m_j' - 1, j - m_j' + 1)$ are also continuation points.

Since m_j is the total number of ones in the first j terms of the sequence $X_1, X_2, \ldots, X_j, \ldots,$ the points $(m_j, j - m_j)$ and $(m_j', j - m_j')$ lie on the line $x + y = j$. Thus our definition 1.5.1 is clearly equivalent to the above definition which states that all the points between any 2 continuation points lying on the line $x + y = j$ are also continuation points.

An alternative and perhaps, a more intuitive way, of characterizing the simplicity of a sampling plan of size n is by means of the number of boundary points. De Groot [6] first proved the interesting result that a sampling plan of size n contains at least $n + 1$ boundary points. From this result and one of Girshick, Mosteller and Savage [9], it follows that a simple sampling plan of size n has exactly $n + 1$ boundary points. This result was obtained in a direct way by Brainerd and Narayana [5].

Mohanty and Narayana [13] used this result to obtain a characterization of simple sampling plans of size n as vectors with $n + 1$ non-negative, integral components satisfying certain conditions. Since we shall use this characterization and some of the results arising from it in later chapters, these results are summarized in this section.

<u>Definition 1.5.2</u> A simple sampling plan is characterized as a vector $a_n = (a_1, a_2, \ldots, a_{n+1})$, the a_i's being non-negative integers satisfying the following conditions:

(a) There exists an integer i, $1 \leq i \leq n$ such that $a_i = a_{i+1} = 0$, i.e. at least 2 consecutive a's in the vector are zero.

(b) Let k be the smallest integer i such that $a_i = a_{i+1} = 0$. Then $a_1 \geq a_2 \geq a_3 \ldots \geq a_{k-1} > 0$ and $0 \leq a_{k+2} \leq a_{k+3} \ldots \leq a_{n+1}$.

(c) Let B be the set of vectors $(b_1, b_2, \ldots, b_k)$ where $b_i = a_i$ $(i = 1,2,\ldots,k-1)$ and let C be the set of vectors $(c_1, c_2, \ldots, c_{n-k})$ where $c_\ell = a_{n+2-\ell}$, $\ell = 1,2,\ldots,n-k$. The b's stand for the a's with indices less than or equal to $k - 1$ and the c's for a's with indices greater than or equal to $k + 2$. Then $b_j \leq n - j$ and $c_\ell \leq n - \ell$. Further, if $b_j = n - j - r$ $(r = 0,1,2,\ldots,n-j-1)$ then

$$c_p \leq \begin{cases} n - p & \text{for } p = 1,2,\ldots,r \\ n - p - j & \text{for } p > r. \end{cases}$$

Similarly, if $c_\ell = n - \ell - r$ $(r = 0,1,\ldots,n-\ell)$, then

$$b_p \leq \begin{cases} n - p & \text{for } p = 1,2,\ldots,r \\ n - p - \ell & \text{for } p > r. \end{cases}$$

It was shown in [13] that to each vector $\underline{a}_n$ there corresponds a simple sampling plan of size n and conversely. Thus, a simple sampling

plan of size n can be represented by a vector of $n + 1$ non-negative, integral components, each component representing the amount of "displacement" of the boundary point from the line $x + y = n$.

Another way of representing a simple sampling plan as a vector was also developed in [13]. It was shown that a simple sampling plan of size n can be represented as a vector $A_{n-1} = (a_1, a_2, \ldots, a_{n-1})$, the a_i's being non-negative integers satisfying the following conditions.

1.5.3

(1) $a_1 \leq a_2 \cdots \leq a_{n-1}$

(2) $a_i \leq 2i \quad i = 1,2,\ldots,n-1.$

In both the characterizations 1.5.2 and 1.5.3 of simple sampling plans of size n, a key role is played by the relation of domination of one sampling plan by another. The concept of domination of vectors with non-negative integral components was first introduced in [15] and has been used in various ways to study many combinatorial problems involving lattice paths. We shall exploit the idea of domination in Chapter 2 to obtain the relationship among the numbers of estimable polynomials and domination with reference to certain sampling plans.

We consider vectors $A_n = (a_1, a_2, \ldots, a_n)$ whose components are non-negative integers satisfying 1.5.3.

Definition 1.5.4 The vector $A_n = (a_1, a_2, \ldots, a_n)$ dominates the vector $B_n = (b_1, b_2, \ldots, b_n)$ (we write $A_n \, d \, B_n$) if, and only if $a_i \geq b_i$ for $i = 1,2,\ldots,n$. We shall say that one sampling plan of size n dominates another sampling plan of size n, if the corresponding vectors dominate each other according to Definition 1.5.4.

Clearly the relationship of domination is a partial order and it can, in fact, be shown that under this relationship the set of vectors A_n, or equivalently the set of all simple sampling plans of size n, form a distributive lattice. We shall have occasion to refer to this idea in Chapter 2.

1.6 Deformations of Simple Sampling Plans.

This section is a summary of the results on deformations of sampling plans introduced in [5] to prove the equivalence of the definitions of simple sampling plans as given by Girshick, Mosteller and Savage [9] and that given by Definition 1.5.2. Since these deformations also play an important role in determining the number of estimable polynomials in a sampling plan, we shall give a brief summary of their properties and define, rather precisely, a special type of deformation called a canonical deformation.

The concept of deformation was defined in [5] for all sampling plans with essential boundaries. Since we are only concerned with the special case of simple sampling plans, we specialize the definition of deformation given in [5] to this case.

<u>Definition 1.6.1</u> Let S be a bounded simple sampling plan. Let $\gamma_o = (x_o, y_o)$ be a boundary point of S. Consider the new sampling plan S' obtained from S as follows:

(i) the boundary of S' consists of the boundary of S except that γ_o becomes a continuation point.

(ii) to the boundary points of S' obtained in (i) add one or both of the points $\gamma_1 = (x_o, y_o + 1)$, $\gamma_2 = (x_o + 1, y_o)$ according as

one or both of the points γ_1, γ_2 were inaccessible in S. Then S' is a deformation of S at the boundary point γ_o. The sampling plan S' is obtained from S by changing one of the boundary points to a continuation point and adding to the boundary one or both of the nearest inaccessible points.

As an illustration of the three possibilities that can occur in Definition 1.6.1, we consider the sampling plan S of size 6, given in the notation of 1.5.2 as (0,0,0,0,1,3,3), (cf. fig. 1). For a deformation at the boundary point $\gamma_o = (2,1)$, (marked ⊕ in fig. 1), replace γ_o by the point (3,1), since the point (2,2) is a continuation point in S. For a deformation of the boundary point $\gamma_o = (3,2)$, (marked X in fig. 1), replace γ_o by the point (4,2), since (3,3) is a boundary point in S. For a deformation at the point $\gamma_o = (3,3)$, we would replace γ_o by the points (3,4) and (4,3), since both of these points are inaccessible in S.

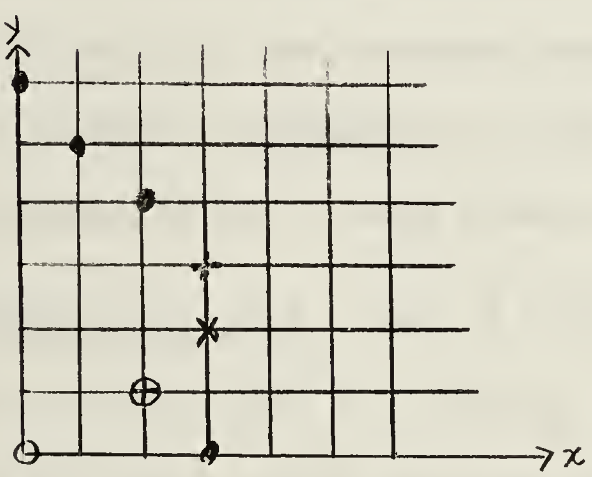


Fig. 1

In considering the above example, it is seen that if we deform at $\gamma_o = (3,0)$, the resulting plan S' would be non-simple. Since we wish to avoid this situation, we make the following definition.

<u>Definition 1.6.2</u> A deformation of S into S' is said to be admissible if S' is simple. In what follows, we shall only consider admissible deformations.

The following theorem, provided in [5], gives the basic result for our work.

Theorem 1.6.1. Let S be any simple sampling plan of size n. Then there exists a sequence of sampling plans $S = S_o, S_1, \ldots, S_k$ such that S_{i+1} is an admissible deformation of S_i $(i = 0,1,2,\ldots,k-1)$ and S_k has exactly the points of index n as boundary points.

There exists a certain sequence of admissible deformations of any simple sampling plan of size n which is of special interest, since it enables us to count the number of right angle turns taken by any path to any boundary point at each stage of the deformation. This sequence, which we call the canonical sequence, enables us to determine the number of estimable polynomials in the basis of the linear space of estimable polynomials for a large class of sampling plans.

Definition 1.6.3 Let $S = (0, \ldots, 0)$ be the "fixed" sampling plan of size n. Let $S' = (a_1, a_2, \ldots, a_{k-1}, 0, \ldots, 0, a_{j+1}, a_{j+2}, \ldots, a_{k+1})$ be any other simple sampling plan of size n, where the a_i's satisfy the conditions of Definition 1.5.2. The sequence of admissible deformations of S' into S given below is called a canonical sequence of deformations.

$$S' = S_o = (a_1, a_2, \ldots, a_{k-1}, 0, \quad 0 \ldots 0, a_{j+1}, a_{j+2} \cdots a_{n+1})$$

$$S_1 = (a_1, a_2, \ldots, a_{k-1}, 0, \quad 0 \ldots 0, a_{j+1}-1, a_{j+2}, \ldots a_{n+1})$$

$$\vdots$$

$$S_{a_{j+1}} = (a_1, a_2, \ldots, a_{k-1}, 0 \ldots 0, 0, a_{j+2}, \ldots, a_{n+1})$$

$$\vdots$$

$$S_{\sum\limits_{j+1}^{n} a_i} = (a_1, a_2, \ldots, a_{k-1}, 0 \ldots 0, a_{n+1})$$

$$S_{1 + \sum\limits_{j+1}^{n} a_i} = (a_1, a_2, \ldots, a_{k-1} - 1, 0 \ldots 0, a_{n+1})$$

$$\vdots$$

$$S_{a_{k-1} + \sum\limits_{j+1}^{n} a_i} = (a_1, a_2, \ldots, a_{k-2}, 0 \ldots 0, a_{n+1})$$

$$\vdots$$

$$S_{\sum\limits_{2}^{k-1} a_r + \sum\limits_{j+1}^{n} a_i} = (a_1, 0 \ldots 0, a_{n+1})$$

$$S_{\sum\limits_{2}^{k-1} a_2 + \sum\limits_{j+1}^{n} a_i + 1} = (a_1, 0 \ldots 0, a_{n+1} - 1)$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$S_{\sum\limits_{2}^{k-1} a_r + \sum\limits_{j+1}^{n+1} a_i} = (a_1, 0 \ldots 0, 0)$$

$$\vdots \qquad \vdots$$

$$S_{\sum\limits_{2}^{k-1} a_r + \sum\limits_{j+1}^{n+1} a_i + a_1 - 1} = (1, 0 \ldots 0, 0)$$

$$S = S_{\sum\limits_{1}^{k-1} a_r + \sum\limits_{j+1}^{n+1} a_i} = (0, 0 \ldots 0, 0)$$

Viewed according to definition 1.6.1, we perform each deformation by moving the "innermost" boundary points, one step at a time, along the lines $x = k - 2, 2, 1, \ldots 0$ and $y = n - j, n - j - 1, \ldots, 1, 0$. Since

at each step of the sequence the components a_i satisfy the conditions of 1.5.2, each step represents an admissible deformation. Moreover, because we are always performing the deformation at an "innermost" boundary point, the situation that arose at the boundary point $(3,0)$ in the example can never occur. Instead of starting with S' and moving the "innermost" boundary points until we reach S, the same sequence can be obtained by starting with S and moving the "outermost" boundary points one step at a time. Thus, this particular sequence of deformations can be performed in reverse order and we write this as the canonical sequence of deformations from S to S'. Since we are primarily interested in what happens as we move from S to S' in this sequence, we shall always refer to the canonical sequence of deformations from S to S', as this causes no confusion.

As an example of canonical sequences and our terminology, consider the canonical sequence for deforming $S = (0, 0, 0, 0, 0, 0)$ into $S' = (2, 2, 0, 0, 1, 2)$.

$$\begin{aligned}
S = S_0 &= (0, 0, 0, 0, 0, 0) \\
S_1 &= (1, 0, 0, 0, 0, 0) \\
S_2 &= (2, 0, 0, 0, 0, 0) \\
S_3 &= (2, 0, 0, 0, 0, 1) \\
S_4 &= (2, 0, 0, 0, 0, 2) \\
S_5 &= (2, 1, 0, 0, 0, 2) \\
S_6 &= (2, 2, 0, 0, 0, 2) \\
S' = S_7 &= (2, 2, 0, 0, 1, 2).
\end{aligned}$$

1.7 Optimal Sampling Plans.

In each of the previous sections, we have been given a binomial sampling plan e.g. the fixed size plan A_n or A_{mn}, and we investigated certain properties pertaining to these plans. In many cases it is useful to determine which sampling plan is "best" to use under certain conditions. Such optimal sampling plans have been described in a very general setting by Blackwell and Girshick [4], Chapters 9 and 10, and a general procedure for finding such plans developed. In a later chapter, we shall investigate the relationship between these optimal plans and the simple sampling plans described in the previous sections. For this purpose, we shall briefly summarize some of the results given in Blackwell and Girshick [4]. The notation used will be that of [4] in general, but for convenience we make some notational changes which shall be indicated. It must be emphasized that this summary is not complete and proofs and other details may be found in [4], Chapters 9 and 10.

Definition 1.7.1 Let $\chi = (X, \Omega, p_w)$ be the sample space where

(i) $X = \{x : x = (x_1, x_2 \ldots x_N)\ x_i = 0 \text{ or } 1, \quad i = 1,2,\ldots,N\}$

(ii) $\Omega \subseteq \Omega' = \{w : 0 \leq w \leq 1\}$

(iii) $p_w(x) = w^{\Sigma x_i} (1 - w)^{N-\Sigma x_i}$

Definition 1.7.2 Let A be any arbitrary space - the space of terminal actions.

Definition 1.7.3 Let $c_j(x)$, $j = 0,1,2,\ldots,N$ be a set of bounded non-negative functions on $J \times X$, where $J = \{0,1,2,\ldots,j\}$ and such that if $x,y \in X$ and $x_i = y_i$, $i = 1,2,\ldots,j$, then $c_j(x) = c_j(y)$.

<u>Definition 1.7.4</u> $L(w,a)$ is a bounded, non-negative function defined on $\Omega \times A$.

<u>Definition 1.7.5</u> Let Ξ be the class of a priori distribution over Ω.

<u>Definition 1.7.6</u> For a fixed $\xi \in \Xi$, and any bounded function h on $\Omega \times X$, let $E_j(h)$ be the conditional expectation of h given $x_1, x_2, \ldots, x_j$, when w has the distribution ξ and for a fixed w, x has the distribution p_w. (Since the ξ in our discussion is always fixed, we suppress the ξ as given in Blackwell and Girshick [4], p. 239).

<u>Definition 1.7.7</u> Let $\varphi_j(x) = \inf_a \ E_j \ L\Big((w,a)\Big)$.

<u>Definition 1.7.8</u> Let $U_j(x) = c_j(x) + \varphi_j(x)$.

<u>Definition 1.7.9</u> Let $\alpha_N(x) = U_N(x)$ and by induction backward define

$$\alpha_j(x) = \min\,[U_j(x),\ E_j(\alpha_{j+1}(x))],\quad j < N.$$

The following theorem is proved in [4].

<u>Theorem 1.7.1</u>. Let ξ be fixed. Let $S^* = (S_0^*, S_1^*, \ldots, S_N^*)$ where $S_j^* = \{x : U_i(x) > \alpha_i(x)$ for $i < j$, $U_j(x) = \alpha_j(x)\}$. The sequential sampling plan S^* is Bayes against ξ, i.e. $\rho(\xi, S^*) = \min_S \rho(\xi, S)$. ($\rho(\xi, S)$ is the risk function).

The optimal sequential procedure is characterized as follows: at the jth stage we compare the present risk $U_j(x)$ with the average risk $E_j(x)$ of going on if we do the best we can with future observations. We stop sampling if $\alpha_j(x) = U_j(x)$, and we take another observation if $\alpha_j(x) = E_j(\alpha_{j+1}(x))$.

The above theorem specifies a general procedure for obtaining optimal sampling plans. In a later chapter, we shall consider the special case of sequential dichotomies for which the cost function is linear and the observations are independent. In this special case, the optimal procedures of Theorem 1.7.1 can be modified in the following way. (cf. [4] sec. 9.3). Rather than consider partitions S^* of the sample space X, we consider regions Ξ_j of the space Ξ (cf. 1.7.5). These regions Ξ_j consist of points ξ_j which are the aposteriori probabilities resulting from ξ and the known probabilities of the first j coordinates of $x \in X$. The optimal procedure is then characterized as follows. Determine the regions $\Xi_0 = \Xi, \Xi_1, \Xi_2, \ldots, \Xi_N$. Given a particular ξ, if $\xi \in \Xi_N$ take no further observations. If $\xi \notin \Xi_N$, a smaller risk will be incurred if we take another observation and follow the optimal procedure S^*_{N-1}. Compute ξ_1 from ξ and the probability distribution of x_1. If $\xi_1 \in \Xi_{N-1}$ stop, otherwise take another observation and compute ξ_2. Proceeding in this manner, we are sure to stop after at most N observations, since $\Xi_0 = \Xi$.

We shall now interpret these results in terms of the binomial sampling plans considered in the previous sections. It is well known that for independent Bernoulli observations the statistic (j, m_j), $m_j = \sum_{i=1}^{j} x_i$, is a sufficient statistic. Thus, we can base our optimal procedure on the sufficient statistics rather than directly on the observations. Since the sets of the partition S^* now are given in terms of m_j, the stopping regions these sets determine also specify the boundary, continuation and inaccessible points in the sampling plans as defined in 1.5. In terms of boundary, continuation and inaccessible points, the

optimal sampling plans can be interpreted in the following manner, where α_j, U_j and E_j are those of definitions 1.7.9, 1.7.8 and 1.7.6 respectively.

Lemma 1.7.1. $\alpha_j(m_j) = E_j\left(\alpha_{j+1}(m_j)\right)$, $j < N$, if and only if $(m_j, j - m_j)$ is a continuation point or an inaccessible point.

Lemma 1.7.2. $\alpha_j(m_j) = U_j(m_j)$, $j \leq N$, if and only if, $(m_j, j - m_j)$ is a boundary point or an inaccessible point.

While the proofs follow directly from the definitions of α_j and those of continuation and boundary points, we have stated these results as lemmas for later reference purposes. Since an inaccessible point can never be reached, it is immaterial whether $\alpha_j = U_j$ or E_j for this point.

CHAPTER II

COMPLETENESS AND ESTIMABLE POLYNOMIALS IN SIMPLE SAMPLING PLANS

2.1 Introduction

In their fundamental paper [9], Girshick, Mosteller and Savage studied unbiased, sequential estimation in one parameter, binomial sampling plans. They showed how to estimate unbiasedly polynomials in the parameter p and gave conditions for the completeness of these sequential plans. These results were extended, among others, by De Groot [6] who proved the following:

1. All polynomials (in p) of degree at most n are estimable unbiasedly for every simple sampling plan of size n.

2. If the boundary of a sampling plan of size n contains more than $n + 1$ points, then the plan is not complete.

3. If the boundary of a sampling plan of size n contains exactly $n + 1$ points, the plan is complete.

We have already investigated completeness and estimation properties of two sampling plans in two parameters in Chapter I. In this Chapter, we shall try to extend De Groot's result to a wider class of two parameter binomial sampling plans and we shall consider certain enumeration problems which arise out of the discussion.

In section 2.3, we prove the completeness of a family of distributions called C_n and list the polynomials estimable unbiasedly. We also examine, in general, two parameter sampling plans, to compare and contrast certain properties of completeness and estimation to those of

one parameter sampling plans considered by De Groot [6]. In particular, we shall prove theorems about two parameter sampling plans analogous to the three theorems of De Groot as described above.

While we are unable to list explicitly all the polynomials estimable unbiasedly in general, in Section 2.3 we determine the dimension of the linear space of polynomials estimable unbiasedly for the class of plans we call "proper" and give an upper bound for any simple sampling plan.

The remaining sections deal with enumeration problems concerned with regular sampling plans which arose in previous sections. We prove some combinatorial theorems analagous to Narayana [18], which enumerate, when specialized, the number of regular sampling plans. We also consider the relationship of domination of sampling plans to the number of estimable polynomials in the plans.

2.2 Completeness and Estimable Polynomials.

Before we proceed to define the sampling plans denoted by C_n, we shall recall certain definitions and prove several results for finite sampling plans.

Let R be the set of continuation points and boundary points determined by a sampling plan of size n. To each boundary point in R, there exist lattice paths which can be thought of as observations on chance variables $X_1, X_2, \ldots, X_n, \ldots,$ where each X_n has the probability distribution defined in 1.2. Thus, to each R there corresponds a sample space $\mathcal{Z} = (Z, \Omega, p)$ where $z \in Z$ represents an observed sequence

of values $X_1, X_2, \ldots,$ $\Omega = \{(0 < p_1 < 1) \times (0 < p_2 < 1),\ p_1 \neq p_2\}$, and $p_w(z)$ represents the probability of a sequence z, given a value $w \in \Omega$. According to 1.2.2, $p_w(z)$ is given by

2.2.1 $$p_w(z) = p_1^{k+1}\, q_1^{y-k}\, p_2^{x-k-1}\, q_2^{k} \quad \text{or}$$

2.2.2 $$p_w(z) = p_1^{k}\, q_1^{y-k}\, p_2^{x-k}\, q_2^{k}$$

depending on whether the last observation of z was a 1 or a 0. k is the number of (1,0) joins and (x,y) is a boundary point of R.

We define a minimal sufficient partition $\mathcal{S}$ on the sample space Z as follows. Let each $s \in S$ consist of all those z for which $p_w(z)$ is constant i.e. if $z_1 \in S$, $z_2 \in S$, then $p_w(z_1) = p_w(z_2)$ for all $w \in \Omega$. On this minimal sufficient partition a sufficient statistic, t, can be defined in such a way that for each $z \in \mathcal{S}$, $t(z)$ will have the same value. The sufficient statistic t will take on as many values as there are sets in the minimal sufficient partition. We need only consider estimators which are functions of the sufficient statistic. Thus, we consider an estimator to be a real-valued function $f(S)$ defined on the partition $\mathcal{S}$. Let $N(S)$ be the number of sequences z in the set $S \in \mathcal{S}$.

<u>Definition 2.2.1</u> Let $P_w(S) = \sum_{z \in S} p_w(z)$. Then $P_w(S) = N(S)\, p_w(z)$, since $p_w(z)$ is constant for each $z \in S$.

<u>Theorem 2.2.1</u> If $\gamma = (x,y)$ is a continuation point or a boundary point in any sampling plan of size n, then any polynomial of the form

(a) $p_1^{k+1}\, q_1^{y-k}\, p_2^{x-k-1}\, q_2^{k}$ $\quad k = 0,1,2,\ldots,\min(x-1,y)$

(b) $p_1^{k}\, q_1^{y-k}\, p_2^{x-k}\, q_2^{k}$ $\quad k = 1,2,\ldots,\min(x,y)$

can be estimated unbiasedly, provided that a path with k (1,0) joins exists.

Proof: We shall prove that a polynomial of type (a) can be estimated unbiasedly. The proof for a polynomial of the type (b) is similar.

First let $\gamma = (x,y)$ be a boundary point and suppose a path with k_o (1,0) joins to (x,y) exists and ends in a one. The probability of such a path is $p_1^{k_o+1}\, q_1^{y-k_o}\, p_2^{x-k_o-1}\, q_2^{k_o}$. This path will represent a sequence $z \in S_j$, for some $S_j \in \mathcal{S}$. Let

$$f(S) = \frac{1}{N(S)} \quad \text{for } S = S_j$$
$$= 0 \quad \text{otherwise.}$$

Then $f(S)$ is the required estimator, for

$$E\Big(f(S)\Big) = \sum_{S \in \mathcal{S}} \frac{1}{N(S)}\, P_w(S)$$
$$= p_1^{k_o+1}\, q_1^{y-k_o}\, p_2^{x-k_o-1}\, q_2^{k_o} .$$

Now suppose $\gamma = (x,y)$ is a continuation point and that a path with k_o (1,0) joins exists to the point γ. Let ' be the partition obtained from the sampling plan, when the origin has been translated to the point $\gamma = (x,y)$. Let us consider all those sequences $z \in$ representing paths which pass through the continuation point (x,y) and have k_o (1,0) joins up to the point (x,y). By assumption, there

exists at least one such sequence. Each such sequence z can be obtained from a sequence $z' \in \mathcal{S}'$ by adjoining to z' a path going to the point $\gamma = (x,y)$ with k_o $(1,0)$ joins. Let us call the class of sets containing the above sequences $\mathcal{S}^{k_o}$. Thus to each set $S' \in \mathcal{S}'$, there corresponds a set $S \in \mathcal{S}^{k_o}$ and conversely. Let $N(S')$ be the number of sequences in S'. For each $S \in S^{k_o}$ let $N'(S) = N(S')$, where S' is the set in $\mathcal{S}'$ corresponding to the set $S \in \mathcal{S}^{k_o}$. Thus, $N(S') = N'(S) \leq N(S)$.

$$\text{Let} \quad f(S) = \frac{N'(S)}{N(S)}, \quad S \in \mathcal{S}^{k_o}$$

$$= 0 \quad \text{otherwise.}$$

Then $f(S)$ is the required estimator, for

$$E\left(f(S)\right) = \sum_{S \in \mathcal{S}} \frac{N'(S)}{N(S)} P_w(S)$$

$$= \sum_{S \in \mathcal{S}} \sum_{z \in S} \frac{N'(S)}{N(S)} N(S) \, p_w(z)$$

$$= \sum_{S \in \mathcal{S}^{k_o}} \sum_{z \in S} N'(S) \, p_w(z)$$

$$= p_1^{k_o+1} q_1^{y-k_o} p_2^{x-k_o-1} q_2^{k_o} \sum_{S' \in \mathcal{S}'} P_w(S')$$

$$= p_1^{k_o+1} q_1^{y-k_o} p_2^{x-k_o-1} q_2^{k_o} .$$

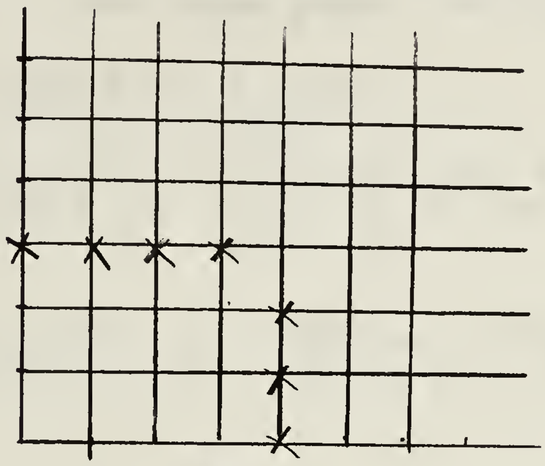
The next theorem is similar to De Groot's Theorem 8.2 [6], which states that if the boundary of a sampling plan of size n contains more than $n + 1$ boundary points then the plan is not complete. The method of proof is similar to that of De Groot and we state the proof very briefly.

Theorem 2.2.2 If the minimal sufficient statistic takes on more than $\frac{(n+1)(n+2)}{2}$ values for any sampling plan of size n, then the plan is not complete.

Proof: $P_w(S)$ is a polynomial in p_1 and p_2 of degree at most n. Hence, the expectation of any estimator is such a polynomial. Thus, the expectation operator maps the space of estimators into the space of polynomials in p_1 and p_2 of degree at most n. Since the dimension of the space of polynomials in p_1 and p_2 of degree at most n is $\frac{(n+1)(n+2)}{2}$, then the dimension of the space of polynomials estimable unbiasedly is $\leq \frac{(n+1)(n+2)}{2}$. Since $\{P_w(S), S \in \mathcal{S}\}$ spans the space of estimable polynomials, then if there are more than $\frac{(n+1)(n+2)}{2}$ polynomials $P_w(S)$, they must be linearly dependent.

In section 1.3 and 1.4, we have already established that not all polynomials in p_1 and p_2 of degree at most n are estimable unbiasedly, since in both the distributions A_n and A_{mn} the polynomials p_2^j, $j = 1,2,\ldots,n$ were not estimable. This situation extends to all finite sampling plans, i.e. in two parameter sampling plans not all the polynomials of degree n in p_1 and p_2 are estimable unbiasedly. Thus, the determination of the polynomials estimable unbiasedly in the two-parameter case is a great deal more complicated than in the one parameter case. With Theorem 2.2.2 and the fact that not all polynomials in p_1 and p_2 of degree n are estimable unbiasedly, we have established the results analagous to the first two theorems of De Groot as stated in the introduction.

We shall illustrate the usefulness of Theorem 2.2.1 by considering a class of sampling plans, C_n. Let C_n denote the simple sampling plans of size n whose boundary consists of the straight lines



$(3,2)_6$

Fig. 2.

(1) parallel to the axes.

(2) a part of the line $x + y = n$ (see Figure 2). Such a sampling plan can be represented by $(a,b)_n$, where a,b are defined as follows:

Among the $n + 1$ boundary points belonging to C_n, let us disregard those that lie on the line $x + y = n$. Of the remaining boundary points, let $a(b)$ form the straight line parallel to the $x(y)$ axis, $a + b \leq n - 1$ $a \geq 0$, $b \geq 0$, i.e. the boundary points not on the line $x + y = n$ lie on the line $y = n - a$ or $x = n - b$. The "fixed" sampling plan of size n is represented by $(0,0)_n$. The plan in Figure 2 is $(3,2)_6$.

To any vector $(a,b)_n$, $a \geq 0$, $b \geq 0$, $a + b \leq n - 1$, there corresponds a plan C_n and conversely i.e. to any 2 composition of the integers $n + 1$, n, ..., 2 corresponds a plan of C_n. The number of r-compositions of n being $\binom{n-1}{r-1}$, the total number of plans in C_n is

$$\binom{n}{1} + \binom{n-1}{1} + \dots + \binom{1}{1} = \binom{n+1}{2}.$$

The coordinates of the boundary points for the sampling plans in C_n are $(0,n-a)$ $(1,n-a)$ ... $(a,n-a)(a + 1, n-a-1)$ ... $(n-b,b)$, $(n-b,b-1)$ ... $(n-b,1)(n-b,0)$ and in the vector notation of 1.5, these plans can be

denoted as

$$(a,\ a-1,\ a-2,\ \ldots,\ 1,\ 0,\ \ldots,\ 0,\ 1,\ 2,\ \ldots,\ b).$$

For these plans, the polynomials $P_w(S)$ are given by the following set, denoted by $S_n(a,b)$.

(a) $\binom{n-a}{k}\binom{r-1}{k-1} p_1^{\,k} q_1^{\,n-a-k} p_2^{\,r-k} q_2^{\,k}$ $\quad r = 1,2,\ldots,a$, $k = 1,2,\ldots,\min(r,n-a)$

(b) $\binom{n-a-j}{k}\binom{a+j-1}{k-1} p_1^{\,k} q_1^{\,n-a-j-k} p_2^{\,a+j-k} q_2^{\,k}$ $\quad j = 1,2,\ldots,n-a-b-1$, $k = 1,2,\ldots,\min(a+j,n-a-j)$

(c) $\binom{n-a-j}{k}\binom{a+j-1}{k} p_1^{\,k+1} q_1^{\,n-a-j-k} p_2^{\,a+j-k-1} q_2^{\,k}$

$j = 1,2,\ldots,n-a-b-1$
$k = 0,1,2,\ldots,\min(a+j-1,n-a-j)$

(d) $\binom{s}{k}\binom{n-b-1}{k} p_1^{\,k+1} q_1^{\,s-k} p_2^{\,n-b-k-1} q_2^{\,k}$ $\quad s = 0,1,2,\ldots,b$, $k = 0,1,2,\ldots,\min(n-b,s)$.

(e) $q_1^{\,n-a}$

The set of polynomials (a) represents paths to the points on the line $y = n-a$, (b) and (c) those paths to the points on the line $x + y = n$, (d) those paths to the points on the line $x = n-b$, and (e) the path to the point $(0,n-a)$. These are polynomials in p_1 and p_2 such that the maximum degree of p_1 is $\max(n-a,b+1)$ and the maximum degree of p_2 is $\max(n-b-1,a)$. A polynomial with the term $p_1^{\,0} p_2^{\,k}$ or $p_1^{\,0} q_2^{\,k}$, $k \neq 0$ cannot be obtained from the above set, since every path starts with probability p_1 or q_1.

Let

2.2.3
$$\max(n-a,\ b+1) = u.$$
$$\max(n-b-1,\ a) = v.$$

Thus, the set $S_n(a,b)$ consists of all polynomials in p_1 and p_2 of maximum degree u in p_1 and v in p_2, $u + v \leq n$.

Lemma 2.2.1 The number of different polynomials in the set $S_n(a,b)$ is $\binom{n+1}{2} - \binom{a+1}{2} - \binom{b+1}{2} + 1.$

Proof: The number of different polynomials in the plan $(0,0)_n$ is $\binom{n+1}{2} + 1$, according to lemma 1.3.2. Starting with the plan $(0,0)_n$, we choose the canonical sequence of deformations that leads to the plan $(a,b)_n$.

$$\begin{array}{l}
(0,\ 0,\ \ldots,\ 0) \\
(1,\ 0,\ \ldots,\ 0) \\
\quad\vdots \\
(a,\ 0,\ \ldots,\ 0) \\
(a,\ 0,\ \ldots,\ 1) \\
\quad\vdots \\
(a,\ 0,\ \ldots,\ 0,\ b) \\
(a,\ 1,\ 0,\ \ldots,\ 0,\ b) \\
\quad\vdots \\
(a,\ a-1,\ a-2,\ \ldots,\ 1,\ 0,\ \ldots,\ 0,\ \ldots,\ b) \\
(a,\ a-1,\ a-2,\ \ldots,\ 1,\ 0,\ \ldots,\ 0,\ \ldots,\ 0,\ 1,\ b) \\
\quad\vdots \\
(a,\ a-1,\ a-2,\ \ldots,\ 1,\ 0,\ \ldots,\ 0,\ 1,\ 2,\ \ldots,\ b).
\end{array}$$

The first deformation in the canonical sequence results in the loss of one polynomial due to the fact that the one path with $k = 0$ to the boundary point $(1, n-1)$ is lost. Similarly the second deformation results in the loss of one path (polynomial) with $k = 0$ to the boundary point $(2, n-2)$. Thus, the first "a" deformations result in the loss of "a" paths. After the first "a" deformations, the next "b" deformation result in the loss of 1 path, corresponding to $k = 1$ to the boundary point $(n-1, 1)$. The next "a-1" deformations result in the loss of 1 path (polynomial) corresponding to $k = 1$ to the boundary point $(2, n-2)$, the next "a-2" deformations result in the loss of 2 polynomials and so forth, until we have performed $\sum_{1}^{a} i + b$ deformations. Similarly, the next "b-1" deformations result in the loss of 2 polynomials corresponding to $k = 1,2$; the next "b-2" deformations result in the loss of 3 polynomials and so forth until after $\sum_{i=1}^{a} i + \sum_{j=1}^{b} j$ deformations, the number lost is $\sum_{i=1}^{a} i + \sum_{j=1}^{b} j$. Thus, the number of polynomials in C_n is

$$\binom{n+1}{2} + 1 - \sum_{i=1}^{a} i - \sum_{j=1}^{b} j = \binom{n+1}{2} - \binom{a+1}{2} - \binom{b+1}{2} + 1.$$

The total number of polynomials can also be obtained by counting the number of polynomials in the set $S_n(a,b)$. The set $S_n(a,b)$ consists of polynomials in p_1 and p_2 of maximum degree u in p_1 and of maximum degree v in p_2 and such that the degree of p_1 plus the degree of p_2 is less than or equal to n. Polynomials of the type $p_1^0 p_2^k$ do not appear. Since $a + b \leq n - 1$, then $n - a \geq b + 1$ and $n - b - 1 \geq a$. Thus, according to 2.2.3, $v = n - b - 1$ and $u = n - a$. Therefore, the

total number of polynomials in $S_n(a,b)$ is $(v + 1)(u + 1) - v - \sum_{i=0}^{n-a-b-1} i$,
where the latter sum represents the number of those polynomials in p_1 and p_2 whose degree is greater than n, since it is readily verified that

$$(v + 1)(u + 1) - v - \sum_{i=0}^{n-a-b-1} i = (n - b)(n - a) + 1 - \frac{(n-a-b-1)(n-a-b)}{2}$$

$$= \binom{n+1}{2} - \binom{a+1}{2} - \binom{b+1}{2} + 1.$$

<u>Theorem 2.2.3</u> The sampling plans C_n are complete.

<u>Proof</u>: By Theorem 2.2.1, we can estimate unbiasedly the following polynomials:

$$\begin{array}{llllll}
1 & p_1 & q_1 p_1 & q_1^2 p_1 & \cdots & q_1^{u-1} p_1 \\
p_1 p_2 & p_1 p_2^2 & p_1 p_2^3 & \cdots & p_1 p_2^v & \\
q_1 p_1 p_2 & q_1 p_1 p_2^2 & q_1 p_1 p_2^3 & \cdots & q_1 p_1 p_2^v & \\
\vdots & & & & & \\
q_1^{u-1} p_1 p_2 & q_1^{u-1} p_1 p_2^2 & \cdots & q_1^{u-1} p_1 p_2^v & &
\end{array}$$

where we omit all polynomials such that the degree of p_1 plus the degree of p_2 is greater than n. The above set of polynomials is linearly independent, contains $(v + 1)u + 1 - \frac{(n-a-b-1)(n-a-b)}{2} = \binom{n+1}{2} - \binom{a+1}{2} - \binom{b+1}{2} + 1$ polynomials and spans the space of estimable polynomials. Since the $\binom{n+1}{2} - \binom{a+1}{2} - \binom{b+1}{2} + 1$ polynomials $P_w(S) \in S_n(a,b)$ also span the space of estimable polynomials, they must be linearly independent. Thus, the theorem is proved.

We finally remark that the class of sampling plans C_n contains the sampling plans A_n and A_{mr} discussed in 1.3 and 1.4, since the plans A_n represent the case where $a + b = 0$ and the plans A_{mr}, the case $a + b = n - 1$. Thus, we have obtained a generalization of the results in these sections. As usual by putting $p_1 = p_2 = p$ in Theorem 2.2.3, we obtain the result that all polynomials in p of degree at most n are estimable unbiasedly, thus once again obtain a generalization of De Groot's Theorem 8.2 [6] to the two parameter case.

Theorem 2.2.1 enables us to give a method for constructing the basis for the linear space of estimable polynomials for a given sampling plan. Consider the set of boundary and continuation points given by the plan. To each of these points $\gamma = (x,y)$ there exists at least one path. Corresponding to this path is a polynomial with degree of $p_1 = y + 1$, (y) and with degree of $p_2 = x - 1$, (x) if the path ends in a horizontal (vertical) step. By Theorems 2.2.1 each such polynomial is estimable. Retain all those polynomials that are of different degree in p_1 or p_2. If the number of polynomials retained is equal to the total number of different polynomials to each boundary point, then we have a basis for the space of estimable polynomials.

Let $\gamma_1, \gamma_2, \ldots, \gamma_r$, $r \geq n + 1$ be the boundary points of a sampling plan of size n. Let $k_1, k_2, \ldots, k_r$ represent the number of different polynomials (paths) to the boundary points $\gamma_1, \gamma_2, \ldots, \gamma_r$. Let Z be the space of polynomials spanned by the $k_1 + k_2 + \ldots + k_r$ polynomials. Then Z is the linear space of estimable polynomials. If the dimension of $Z < k_1 + k_2 + \ldots + k_r$,

then clearly the sampling plan is not complete. On the other hand, if the dimension of $Z = k_1 + k_2 + \ldots + k_r$, then the plan is complete.

We remark that in the case of one parameter sampling plans of size n the dimension of Z is $n + 1$, since every polynomial of at most degree n is estimable unbiasedly. In the two parameter case, not every polynomial in p_1 and p_2 of degree n is estimable unbiasedly and thus the dimension of Z is not easily determined even for simple sampling plans. We present a partial solution to this problem in the next section.

2.3 Upper Bounds for the Number of Estimable Polynomials in the Basis of the Linear Space of Estimable Polynomials in Simple Sampling Plans.

Since, in general, the number of different k values to any particular boundary point determines the total number of values the sufficient statistic takes on, it would be useful if one could determine this number without the tedious process of counting. By using the idea of the canonical sequence of deformations defined in 1.5, we can obtain this number for a large class of simple sampling plans and obtain an upper bound for this number for all simple sampling plans. Since to each path with k (1,0) joins and ending in a vertical or horizontal step there corresponds a polynomial $P_w(S)$ and conversely, we refer to either paths or polynomials since this causes no confusion. By lemma 1.3.2, we know that the number of different polynomials in the "fixed" size sampling plan of size n is $\frac{(n+1)(n+2)}{2} + 1$. By using the canonical sequence of deformations, we can determine how many "k" values are gained or lost at each stage of the sequence i.e. we can count the number of "k" values in each sampling

plan in the canonical sequence. In order to facilitate this counting, we shall classify the sampling plans defined by 1.5.2 into various classes. As a first step, we propose to introduce a particularly simple class of plans called basic plans and then consider two further classes of plans which we shall call "proper" and "regular". In what follows, we shall use the vector notation of definition 1.5.2. We recall that the boundary points denoted by $a_1, a_2, \ldots, a_{k-1}$ in this vector notation refer to displacement parallel to the y axis (i.e. on the left), while the boundary points denoted by $a_{j+1}, \ldots, a_{n+1}$ refer to displacement parallel to the x-axis, (i.e. on the right).

Definition 2.3.1 A sampling plan described by the vector $(a_1, a_2, \ldots, a_{n+1})$ is said to be a basic sampling plan if all the $a_i = 0$ except possibly a_1 or a_{n+1}.

A basic sampling plan has all its boundary points on the line $x + y = n$ except possibly the first and the last. These plans are an intermediate step between any simple sampling plan of size n and the "fixed" size sampling plan.

Definition 2.3.2 A sampling plan described by the vector $(a_1, a_2, \ldots, a_{n+1})$ is said to be regular if the components a_i satisfy the following conditions.

(a') condition (a) of 1.5.2.

(b') Let k be the smallest integer i such that $a_i = a_{i+1} = 0$ and let j, $j \geq k + 1$, be the largest integer i such that $a_i = 0$. Then

$$a_1 > a_2 > \ldots > a_{k-1} > 0 \quad \text{and} \quad a_{j+1} < a_{j+2} < \ldots < a_{n+1}.$$

(c') condition (c) of 1.5.2.

A regular sampling plan is one in which the non-zero components are strictly increasing or decreasing. To put it another way, except for the zero components no two adjacent components are equal. These sampling plans have the property (in the case of the two parameter distributions) that, except for boundary points corresponding to some zero components, only paths that end in a vertical or a horizontal step, but not both, exist to each boundary point corresponding to a non-zero component.

<u>Definition 2.3.3</u> A sampling plan is irregular if it is not regular.

<u>Definition 2.3.4</u> A sampling plan is proper if

(a) $a_j \leq \left\lceil \frac{n+1}{2} \right\rceil - j + 1 \qquad j = 1,2,\ldots,k-1$ and

(b) $a_i \leq \left\lceil \frac{n+1}{2} \right\rceil - (n + 1 - i) \qquad i = k + 2, \ldots, n + 1.$

<u>Definition 2.3.5</u> A sampling plan that is not proper is improper.

A proper sampling plan is one in which all the boundary points not on the line $x + y = n$ with $x \leq \left\lceil \frac{n+1}{2} \right\rceil$ lie on or above the line $y = \left\lceil \frac{n+1}{2} \right\rceil$. Similarly the boundary points with $x \geq \left\lceil \frac{n+1}{2} \right\rceil$ lie on or to the right of the line $x = \left\lceil \frac{n+1}{2} \right\rceil$. In a proper sampling plan, there is no "interference" with the number of paths to any boundary point with $x \leq \left\lceil \frac{n+1}{2} \right\rceil$ by boundary points below it and similarly to any boundary point with $x \geq \left\lceil \frac{n+1}{2} \right\rceil$ by points to the left of it.

With this classification of sampling plans, we can now proceed to determine the number of polynomials in the basis for the linear space of estimable polynomials in each sampling plan. First, we determine this number for basic sampling plans of definition 2.3.1. Then, using the concept

of canonical deformation, we determine this number for all regular, irregular and proper plans. This method also enables us to determine an upper bound for the remaining class of improper plans.

Lemma 2.3.1 The number of different polynomials in a basic plan of size n is given by

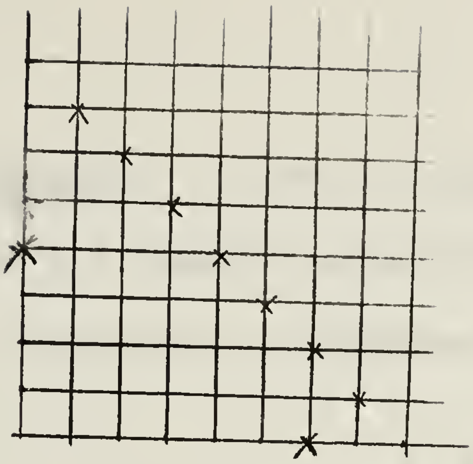
$$(1)\quad E_n = \frac{n(n+1)}{2} + 1 - a_1 \quad \text{if } a_{n+1} = 0$$

$$(2)\quad = \frac{n(n+1)}{2} - a_1 \quad \text{if } a_{n+1} = 1,2,\dots,n-2 \text{ and } a_1 \neq n-1$$

$$(3)\quad = \frac{n(n+1)}{2} + 1 - a_1 - a_{n+1} \quad \text{if } a_{n-1} = n-1 \text{ or } a_1 = n-1.$$

Proof: Let us consider the canonical sequence of deformations from $(0, \dots, 0)$ to the plan $(a_1, 0, \dots, 0, a_{n+1})$. The number of different polynomials in the plan $(0, \dots, 0)$ is $\frac{n(n+1)}{2} + 1$ by Lemma 1.3.1. Consider case (1). The first a_1 deformation moves the boundary point $(0,n)$ to the boundary point $(0,n-a_1)$. Every path to the boundary point $(n-r,r)$, $r > 0$, having $k = 0$ $(1,0)$ joins must pass through the point $(0,r)$ and ends with horizontal steps from $(0,r)$ to $(n-r,r)$. Thus, it is clear that a_1 polynomials have been lost and $E_n = \frac{n(n+1)}{2} + 1 - a_1$. Now consider cases (2) and (3). The first a_1 deformations result in the loss of a_1 polynomials as in case (1). If $a_1 \neq n-1$ and $a_{n+1} = 1,2,\dots,n-2$ one more polynomial is lost; the one corresponding to the path with $k = 1$ $(1,0)$ joins and ending in a vertical step to the point $(n-1,1)$. However, if $a_1 = n-1$ a further $a_{n+1} - 1$ polynomials are lost corresponding to the paths with $k = 1$ $(1,0)$ joins and ending in a vertical step to the points $(n-2,2)$, $(n-3,3)$, $\dots$, $(n-a_n,a_n)$. A similar argument applies if $a_{n+1} = n-1$. Hence the lemma is proved.

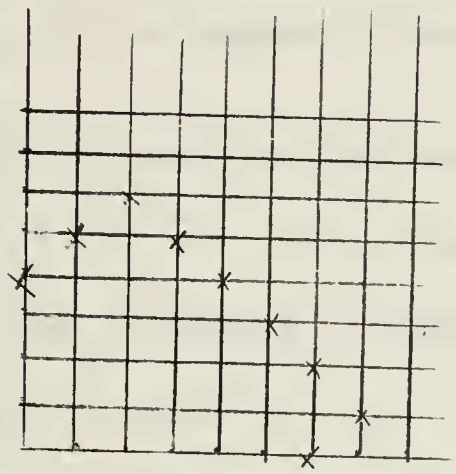
Before we proceed to the next theorem, we shall give an informal explanation of the ideas behind the proof. We recall that, in a regular sampling plan, to all the boundary points corresponding to the components $a_1, a_2, \ldots, a_{k-1}$ there exist only paths that end in a vertical step. Similarly, to all the boundary points corresponding to the components $a_{j+1}, \ldots, a_{n+1}$ only paths that end in a horizontal step exist. From 1.2, we know that the possible k values for paths ending in a vertical step range from 1 to $\min(x,y)$ and for those ending with a horizontal step range from 0 to $\min(x-1,y)$. Therefore, if a plan is proper, we know that $\min(x,y) = x$ and $\min(x-1,y) = x-1$ for all boundary points to the left of the line $x = \left\lceil \frac{n+1}{2} \right\rceil$. Similarly, $\min(x-1,y) = y = \min(x,y)$ for all boundary points to the right of the line $x = \left\lceil \frac{n+1}{2} \right\rceil$. Moreover, in a proper plan, k takes on all the values from 1 to x for all boundary points $\gamma = (x,y)$ to the left of the line $x = \left\lceil \frac{n+1}{2} \right\rceil$, (except for the point $(0,y)$) and all the values from 0 to y for all boundary points to the right of the line $x = \left\lceil \frac{n+1}{2} \right\rceil$ (except for the point $(x,0)$). If we consider a regular, proper plan, it is easily seen that in the canonical sequence of deformations leading to this plan, all the plans are regular and proper. Thus, we need only concern ourselves with the counting of one type of path at each stage of the sequence (ending either in a horizontal or vertical step) and we know how many k $(1,0)$ joins exist for these paths. The following example is given to help clarify these ideas. Consider the canonical sequence of deformations leading to the plan



The Plan S_6

$$S' = (4, 2, 1, 0, 0, 0, 0, 1, 2).$$

$$\begin{aligned}
S = S_0 &= (0, 0, 0, 0, 0, 0, 0, 0, 0) \\
S_1 &= (1, 0, 0, 0, 0, 0, 0, 0, 0) \\
S_2 &= (2, 0, 0, 0, 0, 0, 0, 0, 0) \\
S_3 &= (3, 0, 0, 0, 0, 0, 0, 0, 0) \\
S_4 &= (4, 0, 0, 0, 0, 0, 0, 0, 0) \\
S_5 &= (4, 0, 0, 0, 0, 0, 0, 0, 1) \\
S_6 &= (4, 0, 0, 0, 0, 0, 0, 0, 2) \\
S_7 &= (4, 1, 0, 0, 0, 0, 0, 0, 2) \\
S_8 &= (4, 2, 0, 0, 0, 0, 0, 0, 2) \\
S_9 &= (4, 2, 1, 0, 0, 0, 0, 0, 2) \\
S' = S_{10} &= (4, 2, 1, 0, 0, 0, 0, 1, 2).
\end{aligned}$$



The Plan S_8

Fig. 3

According to lemma 2.3.1, after S_6 the number of polynomials is $\frac{8.9}{2} - 4 = 32$. After S_7, there are 31 since all the paths ending with a horizontal step to the boundary point (2,6) are lost. Since we have already taken into account the paths with $k = 0$ (1,0) joins, we need only count as lost those paths with $k = 1$ (1,0) joins to the point (2,6). After S_8 (i.e. after moving one step), there are still 31, since any further deformations on the line $x = 1$ do not affect any other k values to any other boundary point. In the proof of Theorem 2.3.1, we will utilize the fact that we need count losses of polynomials only at certain stages of the canonical sequence of deformations. Similarly, after S_9, 2 more polynomials are lost and one more is lost after S_{10}. Thus, the total number of polynomials in S' is 27 and this number can easily be verified by actual counting. With this discussion, we can proceed to the theorem.

Theorem 2.3.1 Let $(a_1, a_2, \ldots, a_{k-1}, 0 \ldots 0\ a_{j+1}, \ldots, a_{n+1}) = S'$ represent a regular, proper sampling plan of size n ($j \geq k+1$ and $a_{j+1} \neq 0$). Then, the number of different polynomials in the plan S' is given by

$$A_n = E_n - \left(\sum_{i=1}^{k-2} i + \sum_{\ell=2}^{n-j+1} \ell \right),$$

where E_n is the number of Lemma 2.3.1.

Proof: Let us consider the sequence of canonical deformations from the plan $(0, \ldots, 0)$ to S'. After the first $a_1 + a_{n+1}$ deformations we arrive at a basic plan, according to definition 2.3.1. By lemma 2.3.1, the number of polynomials is given by E_n. The next a_2 deformation result in the loss of 1 polynomial, corresponding to paths with $k = 1$ $(1,0)$ joins and ending in a horizontal step to the point $(3,n-3)$. Similarly, the next a_3 deformations result in the loss of 2 more polynomials corresponding to paths with $k = 1$ and $k = 2$ $(1,0)$ joins and ending in a horizontal step to the point $(4,n-4)$. We proceed in this manner until $\sum_{i=2}^{k-2} a_i$ deformations have been performed losing $3, 4, \ldots, k-2$ polynomials at each stage. Since these plans are regular and proper, no more polynomials are lost. Similarly, after performing the next $\sum_{k=j+1}^{n} a_k$ deformations we lose $2, 3, \ldots, n-j+1$ at each stage corresponding to the k values for all paths ending in a vertical step to the boundary points $(n-3,3) \ldots (n-(n-j+2), n-j+2)$. Since the plans are always regular and proper, no more polynomials are lost, yielding $A_n = E_n - \left(\sum_{i=1}^{k-2} i + \sum_{\ell=2}^{n-j+1} \ell \right)$.

Corollary. If S is a regular, improper sampling plan, the number A_n is an upper bound for the number of polynomials in S.

<u>Proof:</u> If S' is regular and improper at least the number of polynomials as stated in the theorem is lost. In fact, more polynomials can be lost at some stages in the canonical sequence due to the presence of boundary points below or to the right of the boundary points in question. Clearly, A_n is an upper bound.

We now turn to the consideration of irregular sampling plans. We have already seen that in improper plans more polynomials can be lost due to the "interference" of other boundary points. In irregular plans, at certain stages of the canonical sequence, polynomials can be regained because of the "irregular" nature of these plans. Before we prove the next theorem, we shall give an informal discussion of some of the ideas motivating the proof and give several examples to clarify the situation. In the canonical sequence of deformations to a regular plan, all the plans at each stage of the sequence are regular. In regular plans only paths that end in a vertical step exist to boundary points corresponding to the components $a_1, a_2, \ldots, a_{k-1}$ and only paths that end in a horizontal step exist to boundary points corresponding to the components $a_{j+1}, a_{j+2}, \ldots, a_{n+1}$. Moreover, if the plan is proper, then all the plans in the canonical sequence are also proper, and as we have seen in Theorem 2.3.1, only paths ending in a vertical (horizontal) step are lost if the deformations are performed parallel to the $x(y)$ axis. However, if a plan is irregular then both types of path can exist to any boundary point and the simple situation of Theorem 2.3.1 no longer occurs. In fact, some of the paths that were counted as lost may reappear as soon as an irregular plan occurs in the canonical sequence of deformations. We can decide how many paths reappear

by recalling that in performing the last $\sum_{j+1}^{n} a_i$ deformations only paths whose last step is vertical can reappear, while in performing the first $\sum_{2}^{k-1} a_i$ deformations only paths whose last step is horizontal can reappear. If these plans are also proper, all the k values $(k = 1,2,\ldots,\min(x,y))$ corresponding to paths ending in a vertical step can reappear but not all paths ending with a horizontal step may reappear. In particular, paths with $k = 0$ $(1,0)$ joins and ending in a horizontal step will not reappear if there is a boundary point or inaccessible point to the left of the boundary point in question. Thus, if a_i represents the coordinate of the vector at which point an "irregularity" occurs, we can consider the two cases in which

(1) no boundary or inaccessible point lies to the left of the boundary point on the line $y = (n-i+1)-a_i$ and

(2) at least one boundary or inaccessible point lies to the left of the boundary point on the line $y = (n-i+1) - a_i$. Case (1) will occur if $a_r + r < a_i + i$ for all $r = 1,2,\ldots,i-1$ and case (2) will occur if $a_r + r \geq a_i + i$ for at least one r, $r = 1,2,\ldots,i-1$.

<u>Theorem 2.3.2</u> Let $(a_1, a_2, \ldots, a_{k-1}, 0 \ldots 0, a_{j+1}, \ldots, a_{n+1}) = S$ represent an irregular, proper plan of size n. Then the number of different polynomials in S is given by

$$B_n = E_n - \left(\sum_{i=1}^{k-2} i + \sum_{\ell=2}^{n-j+1} \ell\right) + \sum{}'$$

where Σ' is a positive integer determined as follows:

Σ' is obtained by

(i) adding together the indices i, $i = 1,2,\ldots,k-2$ such

that $a_i = a_{i+1}$ and the indices ℓ, $\ell = 1,2,\ldots,j-1$ such that $a_{n-\ell+1} = a_{n-\ell+2}$ and then (ii) subtracting one for each i such that $a_i = a_{i+1}$, $i = 1,2,\ldots,k-2$, if there is at least one r, $r = 1,2,\ldots,i-1$, such that $a_r + r \geq a_i + 1$. Note that in case (1), step (ii) is not necessary. (iii) subtract 1 for each i such that $a_i = a_{i+1}$, if $a_{i+1} + i > a_1$ and for each $r \leq i$, $a_r + r - 1 \not\leq a_{i+1} + i$.

We now give 2 examples to illustrate the computation of Σ'. Consider the irregular, proper plan of size 11 (3,3,3,2,1,0,0,1,1,3,4,4,). Clearly there are no boundary points or inaccessible points to the left of the boundary point corresponding to a_2 and $\Sigma' = 1 + 2 + 1 + 4 - 1 - 1 = 6$. If we take the irregular, proper plan of size 11 (5,3,3,1,1,0,0,1,1,3,4,4), it is easily seen that we are in case (2) and $\Sigma' = 2 + 4 + 1 + 4 - 1 - 1 = 9$.

<u>Proof of Theorem 2.3.2</u>: The fact that we can lose $\left(\sum_{i=1}^{k-2} i + \sum_{\ell=2}^{n-j+1} \ell\right)$ polynomials can be established in the same way as in Theorem 2.3.1. From the discussion preceding the theorem, it is seen that if, at any stage of the canonical sequence an irregular plan is reached, then all the polynomials corresponding to paths ending in a horizontal or vertical step are added in case (1) and in case (2) all of these paths are added except for the paths corresponding to $k = 0$ (1,0) joins. Case (iii) takes care of paths with $k = 0$ (1,0) joins that are lost and not counted by E_n. Thus, the number of polynomials in $\mathcal{S}$ is B_n.

<u>Corollary</u>. The number B_n is an upper bound for the number of polynomials in an irregular, improper sampling plan of size n.

<u>Proof</u>: Because the plan is improper and irregular, at least $\sum_{i=1}^{k-1} i + \sum_{\ell=2}^{n-j+1} \ell$ polynomials are lost and at most Σ' polynomials are added.

2.4 Domination of Sampling Plans and the Number of Estimable Polynomials in the Plan.

In 1.5.3 we defined a sampling plan of size n as the vector $A_{n-1} = (a_1, a_2, \ldots, a_{n-1})$ where the a_i satisfied certain conditions. It can be shown that a regular sampling plan can also be expressed as a vector $A_{n-1} = (a_1, a_2, \ldots, a_{n-1})$, where the a_i are non-negative integers satisfying the following conditions:

2.4.1

(1') Let j be the first integer i such that $a_i > 0$, $i = 1,2,\ldots,n-1$. Then $0 < a_j < a_{j+1} < \ldots < a_{n-1}$.

(2') $a_i \leq 2i$.

It is clear that definition 1.5.4 of domination of vectors applies to the vectors satisfying conditions 2.4.1 above. We seek to establish some relationship between dominated sampling plans and the number of estimable polynomials in such plans. The following theorem establishes such a relationship. Since a rigorous proof of this theorem would require a lengthy rewriting of some previous definitions, we shall given an informal proof indicating how the definition of canonical sequence of deformations can be modified to yield the required result. Let $n(A)$ represent the number of different estimable polynomials in the sampling plan represented by the vector A.

Theorem 2.4.1 If A_n d B_n, and A_n and B_n are both regular, then $n(A_n) \leq n(B_n)$.

Proof: In the definition of the canonical sequence of deformations, we made use of the fact that every sampling plan of size n dominates the

"fixed" sampling plan. We can, in the same way, define a canonical sequence of deformations from any plan B_n to any plan A_n, provided A_n d B_n. Since both A_n and B_n are regular, the canonical sequence of deformations from B_n to A_n yields a regular sampling plan at each stage. Hence, at least a certain number of paths may be lost at each stage and $n(A_n) \leq n(B_n)$.

We now give several examples to illustrate the theorems in this section and section 2.3. Consider $A = (3,2,1,0,0,0,1,3,3)$ according to definition 1.5.2 or equivalently, $A = (0,1,2,3,5,6,6)$ according to definition 1.5.3 and similarly, $B = (3,2,1,0,0,0,1,2,3) = (0,1,2,3,4,5,6)$. B is a regular proper plan and A is an irregular, proper plan. We determine $n(A)$ and $n(B)$ from Theorems 2.3.2 and 2.3.1.

$$n(A) = \left(\frac{8.9}{2}\right) - 3 \;\; - [(1 + 2) + (2 + 3)] + 1 = 26$$

$$n(B) = \left(\frac{8.9}{2}\right) - 3) - [(1 + 2) + (2 + 3)] = 25.$$

It is readily seen that $n(A) - n(B) = 1$, since in A one path ending with a vertical step has been added due to the "irregularity" as the boundary point a_8, $[(4,1)]$. Thus A d B, but $n(A) > n(B)$. Similarly for the irregular proper plans $C = (3,2,1,0,0,0,1,3,3,4)$ and $D = (3,2,1,0,0,0,1,2,3,3)$, we have

$$n(C) = \left(\frac{9.10}{2} - 3\right) - [(1 + 2) + (2 + 3 + 4)] + 2 = 32.$$

$$n(D) = \left(\frac{9.10}{2} - 3\right) - [(1 + 2) + (2 + 3 + 4)] + 1 = 31.$$

C d D but $n(C) > n(D)$. Thus Theorem 2.4.1, which holds only for regular plans, cannot be extended to irregular plans in general.

2.5 Combinatorial Theorems.

In this section, we shall present several theorems which have been mentioned but not fully discussed in the literature [18]. In [18], Narayana proved a combinatorial theorem, analagous to the multinomial theorem, which was used to prove various results concerning vectors with non-negative integral components. This approach yielded, as a special case, the number of simple sampling plans of size n. In this section, we present a theorem analogous to Narayana [18], from which we shall obtain, as a special case, the number of regular sampling plans. We shall also mention very briefly some interesting results that do not pertain directly to regular sampling plans. A complete discussion of the interesting combinatorial aspects of these and other theorems is available in [19].

Theorem 2.5.1 Let n, x_i, δ_i be non-negative integers. Define recursively

$$(0;\ 0 \ldots 0) = 1$$

$$(n;\ x_1,x_2,\ldots,x_k) = 0 \quad \text{if} \quad \sum_{i=1}^{k} x_i > n$$

$$= \sum_{\substack{\delta_i=0 \text{ or } 1 \\ i=1,2,\ldots,k}} (n-1;\ x_1-\delta_1,\ x_2-\delta_2,\ldots,x_k-\delta_k). \quad \text{otherwise.}$$

The sum on the right is taken over 2^k terms. Then

$$(n;\ x_1,x_2,\ldots,x_k) = \prod_{i=1}^{k} \binom{n+1}{x_i} \left[1 - \frac{\sum_{i=1}^{k} x_i}{n+1}\right] \quad \text{if} \quad \sum_{i=1}^{k} x_i \leq n$$

$$= 0 \quad \text{otherwise.}$$

We note that $(n;\ x_1,x_2,\ldots,x_k)$ can be expressed in determinant form as

$$(n;\ x_1,x_2,\ldots,x_k) = \frac{\binom{n+1}{x_1}\cdots\binom{n+1}{x_k}}{(n+1)^k} \begin{vmatrix} n-x_1+1 & -x_2 & \cdots & -x_k \\ -x_1 & n-x_2+1 & \cdots & -x_k \\ \vdots & & & \\ -x_k & \cdots\cdots & & n-x_k+1 \end{vmatrix}$$

<u>Proof</u>: That $(n;\ x_1,x_2,\ldots,x_k) = 0$ if $\sum_{i=1}^{k} x_i > n$ is true by definition. The theorem is true for $n = 0,1$ by direct verification. We assume the theorem is true for all positive integers $n \leq m - 1$ $(m - 1 \geq 1)$. Then, with $\sum_{i=1}^{k} x_i \leq m$, we have

$$(m;\ x_1,x_2,\ldots,x_k) = \sum_{\substack{\delta_i=0 \text{ or } 1 \\ i=1,2,\ldots,k}} (m-1;\ x_1-\delta_1,\ x_2-\delta_2,\ \ldots,\ x_k-\delta_k).$$

We can assume that $\sum_{i=1}^{k} (x_i-\delta_i) \leq m - 1$, since the term containing $\sum_i (x_i-\delta_i) = m$ on the right hand side is by definition zero. Then, by the induction assumption

$$(m;\ x_1,x_2,\ldots,x_k) = \sum_{\substack{\delta_i=0 \text{ or } 1 \\ i=1,2,\ldots,k}} \prod_{i=1}^{k} \binom{m}{x_i-\delta_i} \left(1 - \frac{\sum_{i=1}^{k} (x_i-\delta_i)}{m}\right)$$

$$= \binom{m}{x_1-1} \sum_{\substack{\delta_i=0 \text{ or } 1 \\ i=2,\ldots,k}} \prod_{i=2}^{k} \binom{m}{x_i-\delta_i} \left[1 - \frac{x_1}{m_1} - \frac{\sum_{i=2}^{k} (x_i-\delta_i)}{m} + \frac{1}{m}\right]$$

$$+ \binom{m}{x_1} \sum_{\substack{\delta_i=0 \text{ or } 1 \\ i=2,\ldots,k}} \prod_{i=2}^{k} \binom{m}{x_i-\delta_i} \left(1 - \frac{x_1}{m_1} - \frac{\sum_{i=2}^{k} (x_i-\delta_i)}{m}\right).$$

The two sums on the right consist of 2^{k-1} terms and correspond to putting $\delta_1 = 1$ and $\delta_1 = 0$ in the first expansion. Then $(m; x_1, x_2, \ldots, x_k) =$

$$\left[\sum_{\substack{\delta_i=0 \text{ or } 1\\ i=2,3,\ldots,k}} \prod_{i=2}^{k} \binom{m}{x_i-\delta_i}\left(1 - \frac{x_1}{m} - \frac{\sum_{i=2}^{k}(x_i-\delta_i)}{m}\right)\right]\left[\binom{m}{x_1-1} + \binom{m}{x_1}\right] +$$

$$+ \sum_{\substack{\delta_i=0 \text{ or } 1\\ i=2,3,\ldots,k}} \prod_{i=2}^{k} \binom{m}{x_i-\delta_i}\binom{m}{x_1-1}\frac{1}{m}$$

$$= \binom{m+1}{x_1} \sum_{\substack{\delta_i=0 \text{ or } 1\\ i=2,3,\ldots,k}} \prod_{i=2}^{k} \binom{m}{x_i-\delta_i}\left[1 - \frac{x_1}{m} - \frac{\sum_{i=2}^{k}(x_i-\delta_i)}{m} + \frac{x_1}{m(m+1)}\right]$$

$$= \binom{m+1}{x_1} \sum_{\substack{\delta_i=0 \text{ or } 1\\ i=2,3,\ldots,k}} \prod_{i=2}^{k} \binom{m}{x_i-\delta_i}\left[1 - \frac{x_1}{m+1} - \frac{\sum_{i=2}^{k}(x_i-\delta_i)}{m}\right].$$

Repeating the same argument for $i = 2,3,\ldots,k$ yields the theorem.

Proceeding in a similar way to Theorem 2.5.1, we can easily show that for $\sum x_i \leq n$,

$$\sum_{\delta_i=0 \text{ or } 1} (n; x_1-\delta_1, x_2, x_3, \ldots, x_k) = \binom{n+2}{x_1} \prod_{i=2}^{k} \binom{n+1}{x_i}\left(1 - \frac{x_1}{m+2} - \frac{\sum_{i=2}^{k} x_i}{m+1}\right).$$

We can combine this procedure in the following fashion. Let I_* be a subset $(i_1, i_2, \ldots, i_r)$ of the integers $1,2,\ldots,k$ where $1 \leq i_1 < i_2 < \ldots < i_r \leq k$. Let I denote the complementary set. For $\sum x_i \leq n$, we denote

the sum $\sum_{\delta_{i_1}=0}^{1} \cdots \sum_{\delta_{i_r}=0}^{1} (n;\, x_1,\ldots,x_{i_1}-\delta_{i_1},\, x_{i_2}-\delta_{i_2},\ldots,x_{i_r}-\delta_{i_r},\, x_{i_r+1},\ldots,x_k)$

by $(n;\, x_1,x_2,\ldots,x^*_{i_1}\ldots x^*_{i_r},\, x_{i_r+1},\ldots,x_k)$. By an argument similar to Theorem 2.5.1, we can establish the following theorem.

<u>Theorem 2.5.2</u> For $\Sigma\, x_i \le n$,

$$(n;\, x_1,x_2,\ldots,x^*_{i_1}\ldots x^*_{i_r}\ldots x_k) =$$

$$\prod_{j=1}^{r} \binom{n+2}{x_{i_j}} \prod_{i\,\in\, I} \binom{n+1}{x_i} \left[1 - \frac{\sum_{i_j \in I_*} x_{i_j}}{n+2} - \frac{\sum_{i \in I} x_i}{n+1}\right].$$

From the definition of these numbers, it is obvious that

$$(n;\, x_1\ldots x^*_{i}\ldots x^*_{i_{r-1}},\, x^*_{i_r}\ldots x_k) = (n;\, x_1\ldots x^*_{i_1}\ldots x^*_{i_{r-1}},\, x_{i_r}\ldots x_k)$$
$$+\,(n;\, x_1,\ldots,x^*_{i_1},\ldots,x^*_{i_{r-1}}\;\, x_{i_r}-1,\ldots,x_k).$$

We can associate these numbers with the number of vectors with non-negative integral components satisfying certain conditions. As a special case, we can obtain the number of regular simple sampling plans.

2.6 <u>Non-Negative Increasing Vectors</u>.

Let $a_1,a_2,\ldots$ denote non-negative integers. For $n \ge 1$, let us consider the set of vectors. $A_n = \{a\colon\ a = (a_1,a_2,\ldots,a_n)\}$, when the a_i satisfy the following conditions:

(1) $0 \le a_1 \le a_2 \le \cdots \le a_n$.

2.6.1 (2) $a_i \le ki \quad i = 1,2,\ldots,n$

(3) let j be the first index i, $i = 1,2,\ldots,n$ such that $a_i > 0$. Then $0 < a_j < a_{j+1} < \cdots < a_n$.

Condition (3) ensures that the non-zero components of the vector a are strictly increasing. We note that if we put $k = 2$ in 2.6.1, we obtain the vectors defined by 2.4.1, which can be shown to be in a 1-1 correspondence with the set of regular sampling plans.

Let us now consider the subset $S(n;\ x_1,x_2,\ldots,x_k)$ of A_n, where $S(n;\ x_1,x_2,\ldots,x_k)$ consists of all those vectors in A_n that have exactly x_i of their positive components $\equiv i \pmod k$, $i = 1,2,\ldots,k$, respectively. The following theorem establishes the connection between the number of elements in $S(n;\ x_1,x_2,\ldots,x_k)$ and the numbers $(n;\ x_1,x_2,\ldots,x_k)$ of Theorem 2.5.1.

<u>Theorem 2.6.1</u> The number of vectors in $S(n;\ x_1,x_2,\ldots,x_k)$ is given by the number $(n;\ x_1,x_2,\ldots,x_k)$ of Theorem 2.5.1.

<u>Proof</u>: The theorem is proved by establishing a 1-1 correspondence between the sets $S(n;\ x_1,x_2,\ldots,x_k)$ and the set $T = \bigcup_{\substack{\delta_i=0 \text{ or } 1\\ i=1,2,\ldots,k}} S(n-1;\ x_1-\delta_1, x_2-\delta_2,\ldots,x_k-\delta_k)$. The proof is accomplished by means of the following 1-1 mapping P of S onto T. Let $a \in S(n;\ x_1,x_2,\ldots,x_k)$. Consider the vector $P(a)$ obtained as follows:

(1) replace every element $a_i \leq k$ of a by zero.

(2) replace every element $a_i > k$ by $a_i - k$.

(3) suppress the first zero element of a, leaving a vector of $n-1$ components.

It was shown in [18] that this mapping is a 1-1 mapping of the set $S(n;\ x_1,x_2,\ldots,x_k)$ onto the set $\Sigma = \bigcup_{y_1=0}^{x_1},\ldots,\bigcup_{y_k=0}^{x_k} S(n-1;\ y_1,y_2,\ldots,y_k)$,

where the vectors satisfied only the conditions (1) and (2) of 2.6.1. Since this mapping preserves inequalities, it is readily established that this mapping P is also a 1-1 mapping of $S(n; x_1, x_2, \ldots, x_k)$ onto T.

We can extend the results of Theorem 2.6.1 in the following manner.

<u>Definition 2.6.2</u> Let us consider the set of vectors $B = \{b: b = (b_1, b_2, \ldots, b_n)\}$, where components b_i satisfy the conditions (1) and (3) of 2.6.1 and condition (2) replaced by

(2') $b_i \leq k_i + p$, p an integer such that $0 \leq p \leq k-1$.

Let $S_p(n; x_1, x_2, \ldots, x_k)$ denote the subset of B with the usual congruence properties. Similarly to Theorem 2.6.1, we can prove the following theorem.

<u>Theorem 2.6.2</u> The number of vectors in $S_p(n; x_1, x_2, \ldots, x_k)$ is given by $(n; x_{p+1}, x_{p+2}, \ldots, x_k, x_1^*, x_2^*, \ldots, x_p^*)$, where $(n; x_{p+1}, x_{p+2}, \ldots, x_k, x_1^*, x_2^*, \ldots, x_p^*)$ is the number of Theorem 2.5.2.

2.7 <u>Regular Simple Sampling Plans</u>.

We know that putting $k = 2$ in 2.6.1 yields the vectors that correspond to regular sampling plans. Thus, with the aid of Theorems 2.5.1 and 2.6.1, we can enumerate the regular sampling plans.

Let us define, in general,

$$R(k) = \sum_{c=0}^{n} \sum_{x_1+x_2+\ldots+x_k=c} (n; x_1, x_2, \ldots, x_k)$$

$$= \sum_{c=0}^{n} \left(1 - \frac{c}{n+1}\right) \sum_{x_1+x_2+\ldots+x_k=c} \binom{n+1}{x_1}\binom{n+1}{x_2} \cdots \binom{n+1}{x_k}.$$

Consider the expansion $\left[(1+x)^{n+1}\right]^k = \left[\binom{n+1}{0}x^0 + \binom{n+1}{1}x^1 + \ldots + \binom{n+1}{n+1}x^{n+1}\right]^k$.

It is readily verified that the coefficient of x^c in this expansion is $\sum\limits_{x_1+x_2+\ldots+x_k=c} \binom{n+1}{x_1}\binom{n+1}{x_2}\cdots\binom{n+1}{x_k}$. Since $\left[(1+x)^{n+1}\right]^k = (1+x)^{k(n+1)}$,

then $\sum\limits_{x_1+x_2+\ldots+x_k=c} \binom{n+1}{x_1}\binom{n+1}{x_2}\cdots\binom{n+1}{x_k} = \binom{k(n+1)}{c}$. Thus, $R(k) =$

$\sum\limits_{c=0}^{n} \left(1-\frac{c}{n+1}\right)\binom{k(n+1)}{c}$. For $k = 2$, $R(2) = \sum\limits_{c=0}^{n}\left(1-\frac{c}{n+1}\right)\binom{2(n+1)}{c} = \binom{2n+1}{n}$.

Thus, the number of regular simple sampling plans of size $n+1$ is $\binom{2n+1}{n}$.

For $k = 2$, we can obtain the same results by considering the set of vectors $B_0 = \{b\colon b = (b_1,\ldots,b_n)\}$ as defined in 2.6.2 with $p = 0$ and the set of vectors B_1 as defined in 2.6.2 with $p = 1$. Let us partition the sets B_0 and B_1 into the sets $T_0(n;i)$, $T_1(n;i)$ respectively, $i = 0,1,2,\ldots,n$. Each set $T_j(n;i)$, $j = 0,1$, consists of those vectors in B_0 and B_1 respectively that have exactly i positive components. Let N_0,N_1 represent the number of vectors in B_0,B_1 respectively and $N_0(n;i)$, $N_1(n;i)$ the number of vectors in $T_0(n;i)$ and $T_1(n;i)$ respectively. It can be shown in a manner similar to Theorems 2.5.1 and 2.6.1 that

$$N_0(n;i) = \binom{2n}{i} - \binom{2n}{i-2} \quad \text{and}$$

$$N_1(n;i) = \binom{2n+1}{i} - \binom{2n+1}{i-2}$$

Thus, we can again enumerate the regular, sampling plans, since the total number of plans is $N_0 = \sum\limits_{i=0}^{n} N_0(n;i) = \sum\limits_{i=0}^{n}\left[\binom{2n}{i} - \binom{2n}{i-2}\right] = \binom{2n+1}{n}$.

It is also readily verified that $\binom{2n}{i} - \binom{2n}{i-2} = \left(1-\frac{i}{n+1}\right)\binom{2(n+1)}{i}$.

CHAPTER III

SIMPLICITY IN OPTIMAL SAMPLING PLANS

3.1 Introduction

Girshick, Mosteller and Savage in their fundamental paper, [9], first characterized binomial sampling plans in terms of boundary, continuation and inaccessible points in the plane. In Chapter 1, section 1.5, various characterizations of simple binomial sampling plans have been described. In [4], Blackwell and Girshick gave a description of a more general sampling plan in terms of cylinder sets. In 1.7, we described their procedure for obtaining optimal sampling plans in terms of these cylinder sets. In general, these optimal sampling plans in terms of cylinder sets of Blackwell and Girshick need not necessarily yield the sampling plans that can be characterized by boundary, continuation and inaccessible points in the plane. However, in the cases where the observations are independent, identically distributed Bernoulli random variables, the optimal sampling plans of Blackwell and Girshick do yield the binomial sampling plans as characterized in [9]. It is not at all clear, however, when the optimal sampling plans described by cylinder sets are simple sampling plans in the sense considered in this thesis. It is the purpose of this chapter to present a method by which one can determine whether optimal and other more general sampling plans described by the cylinder sets of [4] are simple sampling plans in the sense considered in this thesis.

Section 3.2 deals with several examples of simple and non-simple optimal plans. In the next section, we present an algorithm by which one can readily determine whether the optimal sampling plans described by the

cylinder sets of Blackwell and Girshick are simple. In the last section, we study the risk functions $E_j(x)$ and $U_j(x)$ in particular problems in order to provide some insight into the structure of simple optimal plans. While it appears possible to formulate certain conditions which would assure simplicity in rather particular cases, we are not able to give any conditions that can be applied in general cases of greater interest.

3.2 Examples of Simple and Non-Simple Optimal Plans.

In this section, we shall give examples of situations that lead to optimal simple sampling plans and an example that leads to an optimal, non-simple sampling plan. The fact that we can obtain a non-simple optimal plan indicates the complete generality of the formulation of these plans as given by Blackwell and Girshick [4]. To determine these optimal sampling plans in particular situations is a formidable, though straightforward, computational problem, as will be evidenced by the examples of this and later sections.

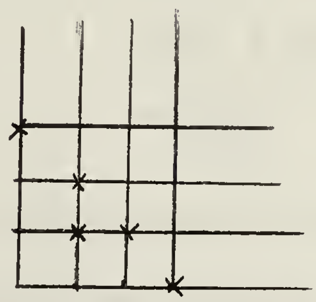
Example 3.2.1 Let $N = 3$, $\Omega = (\frac{1}{4}, \frac{3}{4})$ $A = (\frac{1}{4}, \frac{3}{4})$. Let $\xi(\frac{1}{4}) = \frac{1}{2}$, $\xi(\frac{3}{4}) = \frac{1}{2}$ be the apriori distribution. Let the loss function be $L(\frac{1}{4}, \frac{1}{4}) = 0$, $L(\frac{1}{4}, \frac{3}{4}) = 50$, $L(\frac{3}{4}, \frac{1}{4}) = 100$ and $L(\frac{3}{4}, \frac{3}{4}) = 0$. Let $C_j(x)$ be given by

$$C_0(x) = 0 \qquad C_1(x) = \begin{cases} 1 & \text{if } x_1 = 0 \\ 5 & \text{if } x_1 = 1 \end{cases} \qquad C_2(x) = \begin{cases} 7 & \text{if } \sum_{i=1}^{2} x_i = 0 \text{ or } 2 \\ 3 & \text{if } \sum_{i=1}^{2} x_i = 1 \end{cases}$$

$$C_3(x) = \begin{cases} 3 & \text{if } \sum_{i=1}^{3} x_i = 0 \text{ or } 3 \\ 10 & \text{if } \sum_{i=1}^{3} x_i = 1 \text{ or } 2, \end{cases}$$

for all $x \in X$, when $X = \{x : x = (x_1, x_2, x_3),\ x_i = 0 \text{ or } 1 \text{ for } i = 1,2,3\}$. We denote the value of the conditional expectation $E_j(\alpha_{j+1})$ at m_j by $E_j(m_j)$, cf. 1.7.6. Let $h_j(m_j) = U_j(m_j) - E_j(m_j)$. For this example we exhibit the values of $h_j(m_j)$. $h_2(0) = 1.9$, $h_2(1) = -.75$, $h_2(2) = 1.9$,

$h_1(0) = 5.2$, $h_1(1) = .7$, $h_1(0) = 6.2$. Since $h_2(0) > 0$, $h_2(1) < 0$, $h_2(0) > 0$ and $h_1(0)$, $h_1(1)$ and $h_0(0)$ are all positive, then we can actually determine whether this plan is simple by examining the functions $h_j(m_j)$. In the cases of small N, it is equally convenient to determine the sampling plan specifically in terms of boundary, continuation and inaccessible points in the plan (as illustrated in Figure 4 below).



The Sampling Plan for 3.2.1
Fig. 4

Notice that the cost function in this example is rather unusual in that it depends on x as well as j. Pathological though this example may seem, it is very possible that any general theory will not completely rule out such cases. However, we shall demonstrate explicitly that Blackwell and Girshick [4] have proved the simplicity of optimal sequential sampling plans in the general case of the truncated dichotomy (with a linear cost function and a similar loss function). In most practical cases, the cost function usually satisfies $C_{n+1}(x) \geq C_n(x)$, $n = 0,1,\ldots$ and it seems reasonable to conjecture that the optimal plans are simple in this case.

Example 3.2.2 We now give an example of the more practical kind. Let $N = 5$, $\Omega = \{w : 0 \leq w \leq 1\}$.

$$A = \{a : 0 < a < 1\} \quad \xi(w) = 1 \quad 0 \leq w \leq 1$$
$$= 0 \quad \text{otherwise.}$$
$$L(w,a) = k(w-a)^2 \quad k > 0 \text{ and } C_j(x) = j \text{ for all } x \in X.$$

For this example, the functions $h_j(m_j) = U_j(m_j) - E_j(m_j)$ are symmetric and it can be seen directly from these functions that the optimal plans are simple. We have also determined the actual plans and list them in the following table in terms of the vector notation 1.5.

Table I

k	Sampling Plan
$k > \frac{1764}{5}$	(0, 0, 0, 0, 0, 0)

$\frac{1764}{8} < k \leq \frac{1764}{5}$	(1, 0, 0, 0, 0, 1)
$\frac{1764}{9} < k \leq \frac{1764}{8}$	(2, 1, 0, 0, 1, 2)
$150 < k \leq \frac{1764}{9}$	(1, 0, 0, 0, 1)
$\frac{400}{3} < k \leq 150$	(0, 0, 0, 0).
$100 < k \leq \frac{400}{3}$	(1, 0, 0, 1).
$72 < k \leq 100$	(0, 0, 0)
$36 < k \leq 72$	(0, 0)
$k \leq 36$	No observations required.

We remark that this example is a generalization of an exercise in Blackwell and Girshick [4] p. 245, where $k = 200$. This particular example, which would appear under the third entry for k in Table I is discussed in detail at the end of this chapter, where we hope to provide more information about how the nature of $h_j(m_j)$ determines the simplicity of the plan.

3.3 A Characterization of Sampling Plans in Terms of Basic Cylinder Sets.

Blackwell and Girshick [4], Chapter 3, have defined a sequential sampling plan as a partition $\mathcal{S} = (S_0, S_1, \ldots, S_N)$ of the sample space X, such that each S_j is a cylinder set over $K = \{r \in J;\ 0 < r \leq j\}$, $J = \{0,1,2,\ldots,N\}$. For the sample space $X = \{x : x = (x_1, x_2, \ldots, x_N),\ x_i = 0$ or $1,\ i = 1,\ldots,N\}$ and independent, identically distributed observations x_i, it is difficult to determine whether the optimal sampling plans are simple or not, as seen by the examples in 3.2. In this section, we shall consider a sampling plan as a partition of the sample space by means of the basic cylinder sets, which have already been defined by Blackwell and Girshick [4], P. 239 and we shall give a method by which one can determine whether an optimal plan is simple or not. All of our definitions apply only to the sample space X given above.

<u>Definition 3.3.1</u> For each $x \in X$, let $F_j(x)$ be the set of all points $y \in X$ such that $y \in F_j(x)$ if, and only if, $y_i = x_i$ for $i = 1,2,\ldots,j$. The sets $F_j(x)$ are called basic cylinder sets. In other words, given each $x \in X$, the sets $F_j(x)$ contain all those points in X that have the same first j coordinates as x. Since for each j, there are 2^j possible different first coordinates, there are 2^j different sets $F_j(x)$.

<u>Definition 3.3.2</u> For a fixed j, let B^j be the collection of all sets $F_j(x)$.

Since there are 2^j possible sets in B^j, we can label these sets $B_1^j, B_2^j, \ldots, B_{2^j}^j$. Let $I_r^j = (i_1, i_2, \ldots, i_r)$, $1 \leq i_1 < i_2 < \ldots < i_r \leq 2^j$ be a subset of the set of integers $(1,2,\ldots,2^j)$. As we have already seen, Blackwell and Girshick define a sequential sampling plan as a partition $\mathcal{S} = (S_0, S_1, \ldots, S_N)$ of the sample space X, where each S_j is a cylinder set over the set $K = \{r \in S = 0 \leq r \leq j\}$. Since each $S_j = \bigcup_{i \in I_r^j} B_i^j$, it is readily seen that we can express a sequential sampling plan equivalently in terms of the basic cylinder sets of definition 3.3.1. Thus, we find it convenient to consider a sequential sampling plan as a partition of the sample space X into the basic cylinder sets of definition 3.3.1.

There are $2^j!$ ways of labelling the sets in B^j. We shall choose a particular labelling in order to readily identify the particular sets B_k^j. Let us label the sets in B^j, $j = 0,1,2,\ldots,N$ in the following manner.

<u>Definition 3.3.3</u> B^0 consists of one set, which we label B_1^0. B^1 consists of 2 sets

$$B_1^1 = \{x \in X : x_1 = 0\} \qquad B_2^1 = \{x \in X : x_1 = 1\}.$$

We label the sets B_k^j inductively. Suppose the set B_k^j has already been labelled, $j < N$. Clearly

$$B_k^j = \{x \in X : x \in B_k^j \text{ and } x_{j+1} = 0\} \cup \{x \in X : x \in B_k^j \text{ and } x_{j+1} = 1\}$$

i.e. we partition each $B_k^j \in B^j$ into 2 sets, yielding a total of 2^{j+1} sets. We label every set with $x_{j+1} = 0$ as B_r^{j+1}, where r is to be determined, and similarly every set with $x_{j+1} = 1$ as B_s^{j+1} where s is to be determined. To determine r, s for a particular B_k^j, we proceed as follows:

If $k = 1$, $r = 1$ and $s = 2$. If $k > 1$, k must lie in one of the intervals

$$\left(\sum_{u=0}^{i-1} \binom{j}{u}, \sum_{u=0}^{i} \binom{j}{u} \right] \quad \text{where } i = 1,2,\ldots,j.$$

Then, for

$$k \in \left(\sum_{u=0}^{i-1} \binom{j}{u}, \sum_{u=0}^{i} \binom{j}{u} \right],$$

$$r = k + \sum_{u=0}^{i-1} \binom{j}{u} \quad \text{and} \quad s = k + \sum_{u=0}^{i} \binom{j}{u}.$$

With this labelling the sets B_k^j can be divided into consecutive groups, each group consisting of $\binom{j}{i}$ sets, $i = 0,1,2,\ldots,j$ and such that if $x \in B_k^j$ for any $k \in \left(\sum_{u=0}^{i-1} \binom{j}{u}, \sum_{u=0}^{i} \binom{j}{u} \right]$, then $\sum_{v=1}^{j} x_v = i$. Despite the apparent notational complexity, we illustrate the simple idea behind this labelling with the following example. We indicate the first j coordinate of x underneath each set.

$$B_1^0$$

$$\begin{array}{cc} B_1^1 & B_2^1 \\ 0 & 1 \end{array}$$

$$\begin{array}{cccc} B_1^2 & B_2^2 & B_3^2 & B_4^2 \\ 00 & 01 & 10 & 11 \end{array}$$

$$\begin{array}{cccccccc} B_1^3 & B_2^3 & B_3^3 & B_4^3 & B_5^3 & B_6^3 & B_7^3 & B_8^3 \\ 000 & 001 & 010 & 100 & 011 & 101 & 110 & 111 \end{array}$$

$$\begin{array}{cccccccccccccccc} B_1^4 & B_2^4 & B_3^4 & B_4^4 & B_5^4 & B_6^4 & B_7^4 & B_8^4 & B_9^4 & B_{10}^4 & B_{11}^4 & B_{12}^4 & B_{13}^4 & B_{14}^4 & B_{15}^4 & B_{16}^4 \\ 0000 & 0001 & 0010 & 0100 & 1000 & 0011 & 0101 & 1001 & 0110 & 1010 & 1100 & 0111 & 1011 & 1101 & 1110 & 1111 \end{array}$$

From this example, we see that the sets are labelled so that the sequences with $\sum_{i=1}^{j} x_i = m_j = 0$ appear first, then those with $m_j = 1$ and so forth. For example, in order to label the sets $B_3^4, B_4^4, \ldots, B_8^4$, we first took the 3 sequences corresponding to B_2^3, B_3^3, B_4^3, attached a zero to each sequence and labelled them B_3^4, B_4^4, B_5^4; then we attached a one to each sequence and labelled these sets as B_6^4, B_7^4, B_8^4. Thus, in B^5,

$$B_1^5 = \{x \in X : x_1 = x_2 = x_3 = x_4 = x_5 = 0\}$$

$$B_2^5 = \{x \in X : x_1 = x_2 = x_3 = x_4 = 0 \quad x_5 = 1\}$$

$$B_3^5 = \{x \in X : x_1 = x_2 = x_3 = x_4 = 1, x_5 = 0\}$$ and so forth.

In general, the sequences with the fewest number of ones are labelled first; sequences with the same number of ones are adjacent and can be shown to follow a "dictionary order", although we shall not establish the details.

For each N, we form the triangular array B'_N in the following manner. B'_N consists of $N+1$ rows, the rows being numbered, for the sake of convenience, from 0 to N. The j^{th} row, $j = 0,1,2,\ldots,N$, consists of 2^j terms, where the k^{th} term in the j^{th} row is the B^j_k of definition 3.3.3. In other words, the array B'_N consists of all the basic cylinder sets for the sample space X as ordered by definition 3.3.3. Given any sampling plan $\mathcal{S}$, we can represent it as an array B_N defined below.

<u>Definition 3.3.5</u> Let the sampling plan $\mathcal{S}$ be given. Let B_N be the triangular array $\{B_{jk}\}$ $j = 0,1,2,\ldots,N$, $k = 1,2,\ldots,2^j$ formed from the triangular array B'_N as follows:

(1) retain all the B^j_k given by the sampling plan $\mathcal{S}$, i.e. $B_{jk} = B^j_k$ for $B^j_k \in \mathcal{S}$.

(2) for each $B^j_k \in \mathcal{S}$ retain all the sets B^r_t such that $B^j_k = \bigcup_t B^r_t$, $r = j+1, j+2, \ldots, N$.

(3) replace the remaining B^j_k by zero.

We have, thus, represented a sampling plan $\mathcal{S}$ as an array B_N, consisting of N rows, with 2^j terms in the j^{th} row. Each term in the array is either a zero or an entry B^j_k. With the usual definitions of equality of sampling plans and arrays, it is readily verified that two different sampling plans result in two different arrays B_N. Not every array B_N of zeroes and sets B^j_k will represent a sampling plan. Obviously, one necessary but not sufficient condition is that, with every B^j_k retained, we also retain the sets B^r_t, $r = j+1, j+2, \ldots, N$ such that $B^j_k = \bigcup_t B^r_t$.

As an example of this procedure consider the sampling plan of size 5, $= \{B^2_4, B^3_5, B^3_6, B^4_6, B^4_7, B^4_8, B^5_1, B^5_2, B^5_3, B^5_4, B^5_5, B^5_6, B^5_7, B^5_8, B^5_9, B^5_{10}\}$. The array B_5 is

3.3.1

$$\begin{array}{c}
0 \\
0 \quad 0 \\
0 \quad 0 \quad 0 \quad \underline{B}^2_4 \\
0 \quad 0 \quad 0 \quad 0 \quad \underline{B}^3_5 \quad \underline{B}^3_6 \quad \underline{B}^3_7 \quad \underline{B}^3_8 \\
0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \underline{B}^4_6 \quad \underline{B}^4_7 \quad \underline{B}^4_8 \quad B^4_9 \quad B^4_{10} \quad B^4_{11} \quad B^4_{12} \quad B^4_{13} \quad B^4_{14} \quad B^4_{15} \quad B^4_{16} \\
\underline{B}^5_1 \quad \underline{B}^5_2 \quad \underline{B}^5_3 \quad \underline{B}^5_4 \quad \underline{B}^5_5 \quad \underline{B}^5_6 \quad \underline{B}^5_7 \quad \underline{B}^5_8 \quad \underline{B}^5_9 \quad \underline{B}^5_{10} \quad B^5_{11} \quad B^5_{12} \cdots B^5_{32}
\end{array}$$

The underlined entries are given by $\mathcal{S}$ (rule 1 of definition 3.3.5) and the sets not underlined are obtained from $B^j_k \in \mathcal{S}$ according to rule (2) of 3.3.5.

A sampling plan $\mathcal{S}$, given in terms of the basic cylinder sets above, need not represent a plan of the type that can be described by the boundary, continuation and inaccessible points of Girshick, Mosteller and Savage. To avoid tedious writing, we shall refer to a sampling plan $\mathcal{S}$ given in terms of basic cylinder sets as a BG plan and a sampling plan that can be described in terms of boundary, continuation and inaccessible points as a GSM plan. Every GSM plan can be expressed as a BG plan, but not every BG plan is a GSM plan. A relatively simple solution to the problem of determining when a BG plan is a GSM plan can be given as follows.

<u>Definition 3.3.6</u> Let S be a BG plan. Let the array B_N be given. For each $j = 0,1,2,\ldots,N$, $m_j = 0,1,2,\ldots,j$, let $C_j(m_j)$ be the sequence

$$\left\{B_{jr}, B_{j,r+1}, \ldots, B_{j,r+\binom{j}{m_j}}\right\}, \text{ where}$$

$$r = \sum_{u=o}^{m_j-1} \binom{j}{u} + 1.$$

Definition 3.3.7 Let the sampling plan $\mathcal{S}$ be given. Define the array B_N^0 to be the triangular array obtained from B_N' by applying only rules (1) and (3) of definition 3.3.5 i.e. $B_{jk} = B_k^j$ if $B_k^j \in \mathcal{S}$; $B_{jk} = 0$ otherwise.

For example, the array B_5^0 corresponding to the plan S given by equation 3.3.1 is

$$
\begin{array}{cccccccccccccccc}
 & & & & & & & 0 & & & & & & & & \\
 & & & & & & 0 & & 0 & & & & & & & \\
 & & & & & 0 & 0 & 0 & B_4^2 & & & & & & & \\
 & & & 0 & 0 & 0 & 0 & B_5^3 & B_6^3 & 0 & 0 & & & & & \\
0 & 0 & 0 & 0 & 0 & B_6^4 & B_7^4 & B_8^4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
B_1^5 & B_2^5 & B_3^5 & B_4^5 & B_5^5 & B_6^5 & B_7^5 & B_8^5 & B_9^5 & B_{10}^5 & 0 & 0 & \cdots & 0 & &
\end{array}
\qquad (3.3.2)
$$

Now consider the sequence $C_j(m_j)$, $(m_j \neq 0, j)$. One of the following three possibilities may occur:

(1) $B_{jr} = 0$ for all $B_{jr} \in C_j\,(m_j)$

(2) $B_{jr} = B_r^j$ for all $B_{jr} \in C_j\,(m_j)$

(3) $B_{jr} = 0$ for at least one $B_{jr} \in C_j\,(m_j)$ and
$B_{jS} = B_S^j$ for at least one $B_{jS} \in C_j\,(m_j)$.

We remark that if $m_j = 0$ or j case (3) cannot occur.

Let us consider a BG plan represented by the array B_N. By making the usual identification with lattice paths in the plane, we see that in case (1) every path to the point $(m_j,\ j - m_j)$ exists and continues at least one more step. Thus, this point can be identified as a "continuation" point, since in either the BG or GSM case we continue taking observations (sampling). Similarly, in case (2) we can identify the point $(m_j,\ j - m_j)$ as a "boundary" or "inaccessible" point, since we either continue taking

observations until we reach this point and stop or else we have stopped sampling before. In case (3), we notice the different situations that may arise. Since, there is at least one zero in this sequence, this implies that there is at least one sequence of observations extending beyond this point and thus, this point is a "continuation" point. However, since there is at least one non-zero element, it is possible that the sampling plan $\mathcal{S}$ insist that we take observations until we reach this point and stop. Thus, the point $(m_j, j - m_j)$ could be both a "boundary" and "continuation" point. In a GSM plan such an anomalous situation cannot occur.

The following is an example of a BG plan that is not a GSM plan. $\mathcal{S} = (B_2^2, B_4^2, B_1^3, B_2^3, B_4^3, B_6^3)$. The array B_3 is

3.3.3
$$\begin{array}{ccccccccccccccc} & & & & & & & 0 & & & & & & & \\ & & & & & & 0 & & 0 & & & & & & \\ & & & & & 0 & & B_2^2 & & 0 & & B_4^2 & & & \\ B_1^3 & & B_2^3 & & B_3^3 & & B_4^3 & & B_5^3 & & B_6^3 & & B_7^3 & & B_8^3 \end{array}$$

The array B_3^0 is

3.3.4
$$\begin{array}{ccccccccccccccc} & & & & & & & 0 & & & & & & & \\ & & & & & & 0 & & 0 & & & & & & \\ & & & & & 0 & & B_2^2 & & 0 & & B_4^2 & & & \\ B_1^3 & & B_2^3 & & 0 & & B_4^3 & & 0 & & B_6^3 & & 0 & & 0 \end{array}$$

This sampling plan tells us to take 2 observations and stop if we observe 01 or 11, otherwise take another observation and stop. The following theorem presents a criterion for determining when a BG plan is a GSM plan.

<u>Theorem 3.3.1</u> Let the sampling plan $\mathcal{S}$ be given. Let B_N and B_N^O be the arrays corresponding to S. Let $C_j^O(m_j)$ be the sequence in the array B_N^O corresponding to the sequence $C_j(m_j)$ in B_N. A BG plan is a GSM plan if, and only if, for every sequence $C_j(m_j)$ that has at least one zero the corresponding sequence $C_j^O(m_j)$ has all of its terms zero.

<u>Proof</u>: Suppose a BG plan is a GSM plan. From the discussion preceding, it follows that every point $(m_j, j - m_j)$ corresponding to every sequence $C_j(m_j)$ with at least one zero is a continuation point. In fact, any non-zero terms in $C_j(m_j)$ only appear in B_N by applying rule (2) of definition 3.3.5. But this implies that all the terms in $C_j^O(m_j)$ are zero.

Conversely, suppose to every sequence $C_j(m_j)$ with at least one zero there corresponds a sequence $C_j^O(m_j)$ that has all its terms zero. Then, any non-zero terms appearing in $C_j(m_j)$ appear because of applying rule (2) of definition 3.3.5. Thus, the point $(m_j, j - m_j)$ is a continuation point and the plan is a GSM plan. Thus, the theorem is proved.

Comparing the arrays B_5 and B_5^O given by equations 3.3.1 and 3.3.2, we see that to every sequence $C_j(m_j)$ in B_5 with at least one zero, the corresponding sequence $C_j^O(m_j)$ has all its entries zero. A similar comparison of equations 3.3.3 and 3.3.4 shows that $C_2^O(1)$ corresponding to $C_2(1)$ does not have all its entries zero.

Theorem 3.3.1 provides a criterion for determining when a BG plan is a GSM plan. Once we determine that a plan is a GSM plan, we can then determine whether it is simple or not. From now on, we assume that the sampling plan S is a GSM plan, since we can always determine this fact.

<u>Definition 3.3.8</u> Let the array B_N be given. From the array B_N we form the reduced array B_N^* as follows:

for each $j = 0,1,2,\ldots,N$ and $m_j = 0,1,2,\ldots,j$

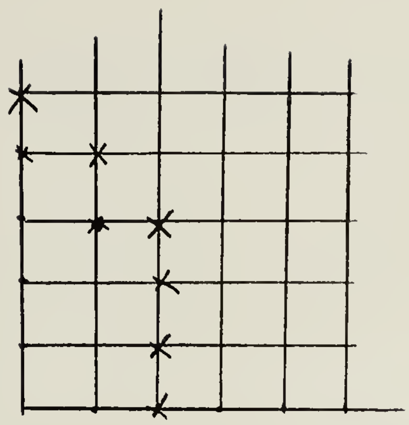
replace the sequence $C_j(m_j)$ by $(B_{m_j}^j)'$ if at least one zero appears in the sequence and by $(B_{m_j}^j)^*$ otherwise. The reduced array B_N^* consists of $N + 1$ rows, the j^{th} row consisting of $j + 1$ terms, each term being distinguished by a prime or a star. From the discussion preceding Theorem 3.3.1, the following lemma follows.

<u>Lemma 3.3.1</u> The point $(m_j, j - m_j)$ is a boundary point or inaccessible point (continuation point) if, and only if $(B_{m_j}^j)^*$ $[(B_{m_j}^j)']$ appears in the reduced array. As an example of this procedure consider the reduced array B_5^* corresponding to the array B_5 given by equation 3.3.1

$$
\begin{array}{cccccc}
 & & (B_0^0)' & & & \\
 & (B_0^1)' & (B_1^1)' & & & \\
 & (B_0^2)' & (B_1^2)' & (B_2^2)^* & & \\
(B_0^3)' & (B_1^3)' & (B_2^3)^* & (B_3^3)^* & & \\
(B_0^4)' & (B_2^4)' & (B_2^4)^* & (B_3^4)^* & (B_4^4)^* & \\
(B_0^5)^* & (B_1^5)^* & (B_2^5)^* & (B_3^5)^* & (B_4^5)^* & (B_5^5)^*.
\end{array}
$$

This array represents the sampling plan illustrated in Figure 5. In terms of the vector notation, this is the plan (0, 0, 0, 1, 2, 3)

In the reduced array, the primed sets appear in runs, in the usual sense, in each row. Thus, we can use the reduced array to determine



The Sampling Plan for 3.3.1

Figure 5.

whether a GSM sampling plan is simple.

Theorem 3.3.2 The GSM sampling plan is simple if, and only if, the maximum number of runs of primed sets in each row is one.

Proof: S is simple if, and only if, there are no boundary or inaccessible points between the continuation points on the line $x + y = j$, $j = 1,2,\ldots,N-1$. Because of lemma 3.3.1, for each j we can identify the non-continuation and continuation points with the starred and primed sets respectively. Thus, the above statement is equivalent to the theorem.

As an example of a non-simple GSM plan consider

$$\mathcal{S} = (B_2^2,\ B_3^2,\ B_2^3,\ B_8^3,\ B_1^4,\ B_2^4,\ B_{11}^4,\ B_{14}^4).$$

B_4 is

$$\begin{array}{cccccccc} & & & & 0 & & & \\ & & & 0 & & 0 & & \\ & & 0 & B_2^2 & & B_3^2 & 0 & \\ 0 & B_2^3 & B_3^3 & B_4^3 & B_6^3 & B_6^3 & 0 & B_8^3 \\ B_1^4 & B_2^4 & \ldots & & & & \ldots & B_{16}^4 . \end{array}$$

The reduced array B_4^* is

$$
\begin{array}{ccccccccc}
 & & & & (B_0^0)' & & & & \\
 & & & (B_0^1)' & & (B_1^1)' & & & \\
 & & (B_1^2)' & & (B_1^2)^* & & (B_2^2)' & & \\
 & (B_0^3)' & & (B_1^3)^* & & (B_2^3)' & & (B_3^3)^* & \\
(B_0^4)^* & & (B_1^4)^* & & (B_2^4)^* & & (B_3^4)^* & & (B_4^4)^* .
\end{array}
$$

This sampling plan is not simple because in rows 2 and 3 there is more than one run of primed sets. This plan is illustrated in Figure 6.

We finally remark that we can determine whether a plan is a GSM plan by comparing the array B_N directly with the sampling plan $\mathcal{S}$, without explicitly introducing the array B_N^0. However, for sake of exposition and clarity, we found it necessary to define the arrays B_N^0. Let $\mathcal{S}$ be a BG plan given in terms of the sets B_k^j. Let B_N be the array corresponding to S. We examine all the sequences $C_j(m_j)$ in B_N that contain at least one zero. If, in any such sequence $C_j(m_j)$ any terms B_{jr} apear that are given directly by S, the sampling plan is not a GSM plan. In a similar way, we can easily determine whether a GSM plan is simple directly from the array B_N without explicitly introducing the arrays B_N^*, although we shall not go into the details here. Thus, given a sampling plan $\mathcal{S}$ in terms of the cylinder sets of Blackwell and Girshick, we can determine whether it is a plan that can be described by the boundary, continuation and inaccessible points of Girshick, Mosteller and Savage and then decide whether the plan is simple.

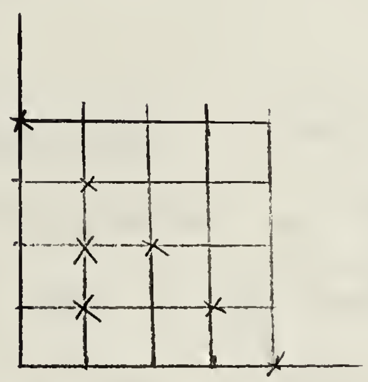


Figure 6.

3.4 Optimal Sampling Plans and Simplicity.

From the discussion of optimal sampling plans in 1.7, we see that an optimal plan is determined by the risk functions E_j and U_j for each j and m_j. Thus, it is reasonable to expect that the property of simplicity of optimal sampling plans will, in some way, depend upon the behaviour of these risk functions.

We now restate the condition for simplicity of optimal sampling plans in terms of the risk functions E_j and U_j.

Definition 3.4.1 For each $j = 0,1,2,\ldots,N-1$ we let $h_j(m_j) = U_j(m_j) - E_j(m_j)$.

Definition 3.4.2 Let m_j' be the least m_j for which $h_j(m_j) > 0$ and m_j'' be the greatest m_j for which $h_j(m_j) > 0$, $0 \leq m_j \leq j$. We remark that for some j, m_j' and m_j'' may be equal or $h_j(m_j) \leq 0$ for all m_j, so that m_j' and m_j'' are undefined.

From lemmas 1.7.1 and 1.7.2, it follows that for a boundary or inaccessible point, $h_j(m_j) \leq 0$ and for a continuation point or inaccessible point, $h_j(m_j) > 0$.

Thus, the condition for simplicity can be stated as follows.

Lemma 3.4.1 If for every $j = 0,1,2,\ldots,N-1$, $h_j(m_j) > 0$ for all m_j such that $m_j' \leq m_j \leq m_j''$, then the optimal plan is simple.

We mention 2 classes of functions that satisfy the conditions of lemma 3.4.1:

Condition (1) $h_j(m_j)$ is a monotonic function of m_j.

Condition (2) $h_j(m_j)$ is monotonic increasing for $m_j \leq x$, and

monotonic decreasing for $m_j \geq x$, for some x such that $0 < x < j$. Unfortunately, the determination of the functions $h_j(m_j)$ is a complicated computational procedure. Also, the determination of these functions is essentially equivalent to determining the sampling plan. A more meaningful question to ask would be what restrictions or conditions must we place on the loss functions, cost functions, apriori distributions and so on, to guarantee that the functions $h_j(m_j)$ will satisfy the conditions for simplicity or more directly, what classes of loss functions, cost functions, apriori distributions will result in simple sampling plans. Although we are unable to give such general conditions, we do give one class of situations which always yield simple optimal plans. This class consists of the truncated sequential dichotomy considered in Blackwell and Girshick [4] and the proof of the simplicity of the optimal plans for this case follows directly from their results as shown in Example 3.4.1 below. We also give a numerical example, where it can be verified that the functions $h_j(m_j)$ satisfy the conditions for simplicity without actually computing all these functions. We conjecture that in most of the usual "practical" problems, the optimal sampling plan is simple.

Example 3.4.1 The Truncated Sequential Dichotomy. Let

$$\Omega = (w_1, w_2) \quad w_1 \neq w_2, \qquad A = (a_1, a_2).$$

We assume $0 < w_1 < 1$ and $0 < w_2 < 1$. Let $\xi(w_1) = \xi_1$ and $\xi(w_2) = 1 - \xi_1$, $0 < \xi_1 < 1$, be the apriori distribution over Ω. Let $c_j(x) = j$, $j = 0,1,2,\ldots,N$. $L(w, a)$ is given by the following matrix.

$$\begin{array}{cc} & \begin{array}{cc} a_1 & a_2 \end{array} \\ \begin{array}{c} w_1 \\ w_2 \end{array} & \begin{bmatrix} 0 & b_{12} \\ b_{21} & 0 \end{bmatrix} \end{array} \qquad \text{where } b_{12} > 0 \text{ and } b_{21} > 0.$$

Then $\xi_j^1(m_j)$ = conditional probability of w_1 given $x_1, x_2, \ldots, x_j$

$$= \frac{\xi_1 w_1^{m_j} (1 - w_1)^{j-m_j}}{\xi_1 w_1^{m_j} (1 - w_1)^{j-m_j} + (1 - \xi_1) w_2^{m_j} (1 - w_2)^{j-m_j}}$$

$$\xi_j^2 (m_j) = 1 - \xi_j^1(m_j).$$

In 3.2, we have given an example of a non-simple optimal plan with a similar loss function but a non-linear cost function. It follows readily from the following results in Blackwell and Girshick [4] p. 263 that with a linear cost function, we always obtain a simple, optimal plan.

<u>Theorem 3.4.1</u> (Blackwell and Girshick) There exist numbers γ_{N-j} and δ_{N-j}, $j = 0,1,2,\ldots,N$ such that if

(1) $\delta_{N-j} \leq \xi_j^1 \leq 1$ we choose a_1 without further observations.

(2) $0 \leq \xi_j^1 \leq \gamma_{N-j}$ we choose a_2 without further observations.

(3) $\gamma_{N-j} < \xi_j^1 < \delta_{N-j}$ we take another observation.

The important fact that enables us to show simplicity is that $\xi_j^1(m_j)$ is a monotonic function of m_j, (for a fixed j).

$$\xi_j^1(m_j) = \frac{1}{1 + \frac{1 - \xi_1}{\xi_1} \left(\frac{1 - w_2}{1 - w_1}\right)^j \left(\frac{w_2(1 - w_1)}{w_1(1 - w_2)}\right)^{m_j}}$$

$$= \frac{1}{1 + K \left[\frac{w_2(1 - w_1)}{w_1(1 - w_2)}\right]^{m_j}}$$

If $w_2 < w_1$, then $\frac{w_2(1 - w_1)}{w_1(1 - w_2)} < 1$ and $\xi_j^1(m_j)$ is increasing with m_j. If

$w_1 < w_2$, then $\xi_j'(m_j)$ is decreasing with m_j.

We now assume that $\xi_j'(m_j)$ is increasing with m_j. A similar proof holds if $\xi_j'(m_j)$ is decreasing with m_j. We can also assume that $\xi_1 \in (\gamma_N, \delta_N)$, for otherwise we would take no observations and the plan is trivially simple. Consider $\xi_1'(0)$ and $\xi_1'(1)$. Either

(a) $\xi_1'(0) \in [0, \gamma_{N-1}]$

(b) $\xi_1'(0) \in [\gamma_{N-1}, \delta_{N-1}]$ or

(c) $\xi_1'(0) \in [\delta_{N-1}, 1]$. Since $\xi_1'(1) > \xi_1'(0)$, then $\xi_1'(1)$ can lie in one of the following 3 intervals corresponding to the three cases (a), (b) and (c)

(a') $\xi_1'(1) \in [0,1]$

(b') $\xi_1'(1) \in (\gamma_{N-1},1]$

(c') $\xi_1'(1) \in [\delta_{N-1},1]$.

Thus, either the points $(1,0)$ and $(0,1)$ are both boundary points or both continuation points or one is a boundary point and one a continuation point. We now assume that we have continued sampling to $j = k$ observations, that all the points on the line $x + y = r$, $r \leq k$ satisfy the conditions for simplicity and that the points $(m_k', k - m_k') \ldots (m_k' + h, k - m_k' - h)$ $h < k - m_k'$ are continuation points and all the other points on the line $x + y = k$ are either boundary points or inaccessible. Then since $\xi_{k+1}'(m_k + h + 1) > \xi_{k+1}'(m_k')$, according to Theorem 3.4.1 only the following cases can arise.

(1) $\xi_{k+1}'(m_k') \in (\gamma_{N-k-1}, \delta_{N-k-1})$ and $\xi_{k+1}'(m_k' + h + 1) \in (\gamma_{N-k-1}, \delta_{N-k-1})$

(2) $\xi_{k+1}'(m_k') \in (\gamma_{N-k-1}, \delta_{N-k-1})$ and $\xi_{k+1}'(m_k' + h + 1) \in [\delta_{N-k-1}, 1]$.

(3) $\xi'_{k+1}(m'_k) \in [0, \gamma_{N-k-1}]$ and $\xi'_{k+1}(m'_k + h + 1) \in (\gamma_{N-k-1}, \delta_{N-k-1})$.

(4) $\xi'_{k+1}(m'_k) \in [0, \gamma_{N-k-1}]$ and $\xi'_{k+1}(m'_k + h + 1) \in [\delta_{N-k-1}, 1]$.

(5) $\xi'_{k+1}(m'_k) \in [0, \gamma_{N-k-1}]$ and $\xi'_{k+1}(m'_k + h + 1) \in [0, \gamma_{N-k-1}]$.

(6) $\xi'_{k+1}(m'_k) \in [\delta_{N-k-1}, 1]$ and $\xi'_{k+1}(m'_k + h + 1) \in [\delta_{N-k-1}, 1]$.

(Note that these 6 cases are exactly analagous to the cases (a), (b), (c), (a'), (b'), (c') considered simultaneously, where naturally $m'_k = 0$ and $h = 0$). We remark that if there exist any points for which $m_{k+1} < m'_k$ and $m_{k+1} > m'_k + h + 1$, then these could only be boundary or inaccessible points. Since $\xi'_{k+1}(m_{k+1})$ is an increasing function of m_{k+1}, we see that in the first 4 cases all the continuation points on the line $x + y = k + 1$ form an interval. For example, in case 3, since $\xi_{k+1}(m'_k) \in [0, \gamma_{N-k-1}]$ and $\xi'_{k+1}(m'_k + h + 1) \in (\gamma_{N-k-1}, \delta_{N-k-1})$, there is a first value m''_{k+1} such that $m'_k < m''_{k+1} \leq m'_k + h + 1$ and $\xi'_{k+1}(m''_{k+1}) \in (\gamma_{N-k-1}, \delta_{N-k-1})$. Thus, all the points such that $m_{k+1} < m''_k$ and $m_{k+1} > m''_k + h + 1$ are boundary or inaccessible points, while all the points such that $m''_{k+1} \leq m_{k+1} \leq m''_k + h + 1$ are continuation points. Similarly in case (5) and (6) all the points on the line $x + y = k + 1$ are boundary or inaccessible points. Thus, the sequential dichotomy yields a simple sampling plan.

<u>Example 3.4.2</u> We illustrate with a particular numerical example that we can establish that $h_j(m_j)$ satisfies the conditions of Lemma 3.4.1 without actually calculating $h_j(m_j)$ for all j. In this case also the functions $h_j(m_j)$ satisfy condition (2) of Lemma 3.4.1.

Let $\Omega = (w, 0 \leq w \leq 1)$ $\qquad$ $A = (a : 0 \leq a \leq 1)$.

Let $\xi(w) = 1$, $\quad c_j(x) = j$ and $L(w,a) = k(w - a)^2 \quad k > 0$.

Then $\xi_j(w) = \dfrac{w^{m_j}(1-w)^{j-m_j}}{\int_0^1 w^{m_j}(1-w)^{j-m_j}\,dw}$.

From definition 1.7.8

3.4.2 (1) $$U_j(m_j) = j + k \cdot \min_a \int_0^1 (w-a)^2 \xi_j(w)\,dw$$
$$= j + \frac{k(1+m_j)(j+1-m_j)}{(j+2)^2(j+3)}$$

We denote the particular value of the conditional expectation $E_j(U_{j+1})$ by $E_j^U(m_j)$ and similarly, particular values of $E_j(\alpha_{j+1})$ by $E_j(m_j)$ (c.f. 1.7.6 P. 25). Then

3.4.2 (2) $$E_j^U(m_j) = j + 1 + \frac{k}{(j+3)^2(j+4)}\Big\{(1+m_j)(j+1-m_j) + (j-2m_j)\,E_j(x_{j+1}) - E_j(x_{j+1}^2)\Big\}$$
$$= j + 1 + \frac{k}{(j+2)(j+3)^2}(1+m_j)(j+1-m_j),$$

where we have used the fact that $E_j(x_{j+1}) = E_j(x_{j+1}^2) = \dfrac{m_{j+1}}{j+2}$. Now, by definition $h_j(m_j) = U_j(m_j) - E_j(m_j)$ and we let $g_j(m_j) = U_j(m_j) - E_j^U(m_j)$. It can be easily shown from 3.4.2 (1) and 3.4.2 (2) that

$$g_j(m_j) = -1 + \frac{k}{(j+2)^2(j+3)^2}(1+m_j)(j+1-m_j).$$

Since $E_j(\alpha_{j+1}) = E_j\,[\min(U_{j+1}, E_{j+1})]$, then it follows that $h_j(m_j) \geq g_j(m_j)$ for all j and m_j.

We now consider the case when $N = 5$ and $k = 200$ to illustrate how these results can be used to establish that this plan is simple: (cf. 3.2)

Now $h_4(m_j) = g_4(m_j)$. For $k = 200$,

$$g_4(2) = -1 + \frac{200(3)(3)}{6^2 \cdot 7^2} > 0 \quad \text{and} \quad g_4(1) = -1 + \frac{200(2)(4)}{6^2 \cdot 7^2} < 0.$$

Since $g_j(m_j)$ is symmetric and monotonically increasing for $m_j \leq j/2$, we can conclude that $g_4(0) < 0$ and $g_4(3) < 0$, $g_4(4) < 0$. Similarly, we can establish that $g_3(0) < 0$, $g_3(1) > 0$, $g_3(2) > 0$, $g_3(3) < 0$, and $g_2(m_2) > 0$ for $m_2 = 0,1,2$, $g_1(m_1) > 0$ for $m_1 = 0,1$ and $g_0(0) > 0$. Thus, $h_4(m_4) > 0$ for $m_4 = 2$, $h_4(m_4) < 0$ for $m_4 = 0,1,3,4$. Similarly, since $h_j(m_j) \geq g_j(m_j)$, $h_3(m_3) > 0$ for $m_3 = 1,2$, $h_3(m_3) < 0$ for $m_3 = 0,3$; $h_2(m_2) > 0$ for all m_2, $h_1(m_1) > 0$ for all m_1. Thus, $h_j(m_j)$ satisfies condition (2) of Lemma 3.4.1 and the plan is simple.

Although a similar procedure could be carried out for any N and k, a detailed description taking care of all the possible cases that could arise is formidable computational problem. In all the numerical examples we have attempted, all the optimal plans were simple and the functions $h_j(m_j)$ satisfied condition (2) of Lemma 3.4.1, and were symmetric about the point $x = j/2$, supporting the conjecture that all optimal plans for this situation are simple. However, our methods are not powerful enough to establish this conjecture.

REFERENCES

[1] BAHADUR, R.R., "Sufficiency and Statistical Decision Functions", The Annals of Mathematical Statistics, XXV (1954), p. 423.

[2] BAHADUR, R.R., "On Unbiased Estimated of Uniformly Minimum Variance", Sankhya, XVIII (1957), p. 211.

[3] BLACKWELL, D., "Conditional Expectation and Unbiased Sequential Estimation", The Annals of Mathematical Statistics, XVIII (1947), p. 105.

[4] BLACKWELL, D. and GIRSHICK, M.A., "Theory of Games and Statistical Decisions", John Wiley and Sons, Inc., New York, 1954.

[5] BRAINERD, B. and NARAYANA, T.V., "A Note on Simple Binomial Sampling Plans", The Annals of Mathematical Statistics, XXXII (1961), p. 906.

[6] DE GROOT, M.H., "Unbiased Sequential Estimation for Binomial Populations", The Annals of Mathematical Statistics, XXX (1959), p. 80.

[7] FELLER, W., "An Introduction to the Theory of Probability and Its Applications", Vol. I, John Wiley and Sons, Inc., New York, 1957.

[8] GABRIEL, K.R., "The Distribution of the Number of Successes in a Sequence of Dependent Trials", Biometrika. 46, 1959. p. 454.

[9] GIRSHICK, M.A., MOSTELLER, F. and SAVAGE, L.J., "Unbiased Estimates for Certain Binomial Sampling Plans with Applications", The Annals of Mathematical Statistics, XVII (1946), p. 13.

[10] LEHMANN, E.L. and SCHEFFE, H., "Completeness, Similar Regions and Unbiased Estimation", Part I, Sankhya, X (1950), p. 305.

[11] LEHMANN, E.L. and STEIN, C., "Completeness in the Sequential Case", The Annals of Mathematical Statistics, XXI (1950), p. 376.

[12] MOHANTY, S.G. and NARAYANA, T.V., "Some Properties of Compositions and their Applications to Probability and Statistics I", Biometrische Zeitschrift, III (1961), p. 252.

[13] MOHANTY, S.G. and NARAYANA, T.V., "Some Properties of Compositions and their Applications to Probability and Statistics II", Biometrische Zeitschrift, V (1963), p. 8.

[14] NARAYANA, T.V., "Sur Les Treillis Formés Par Les Partitions d'un Entier et Leurs Applications à la Theorie des Probabilités", Comptes Rendus, CCXL (1955), p. 1188.

[15] NARAYANA, T.V. and FULTON, G.E., "A Note on the Compositions of an Integer", Canadian Mathematical Bulletin, I (1958), p. 169.

[16] NARAYAYA, T.V., "A Partial Order and its Applications to Probability Theory", Sankhya, XXI (1959), p. 91.

[17] NARAYANA, T.V. and SATHE, Y.S., "Minimum Variance Unbiased Estimation in Coin Tossing Problems", Sankhya, XXIII (1961), p. 183.

[18] NARAYANA, T.V., "An Analogue of the Multinomial Theorem", Canadian Mathematical Bulletin, V (1962), p. 43.

[19] ROHATGI, V.K., "On Some Combinatorial Theorems Arising in Probability and Statistics", Unpublished Master's Thesis.

[20] SAVAGE, L.J., "A Uniqueness Theorem for Unbiased Sequential Estimation", The Annals of Mathematical Statistics, XVIII (1947), p. 295.

[21] WOLFOWITZ, J., "On Sequential Binomial Estimation", The Annals of Mathematical Statistics, XVII, p. 409.